

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**



(51) International Patent Classification <sup>6</sup> :		(11) International Publication Number:	
A61K 38/00, 38/02, C07K 5/00, 7/00		A2	(43) International Publication Date: 19 November 1998 (19.11.98)
(21) International Application Number:			PCT/US98/10088
(22) International Filing Date:			15 May 1998 (15.05.98)
(30) Priority Data:			
60/046,595	15 May 1997 (15.05.97)	US	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application			
US	60/046,595 (CIP)		
Filed on	15 May 1997 (15.05.97)		
(71) Applicants (for all designated States except US): CYTOGEN CORPORATION [US/US]; 600 College Road East, Princeton, NJ 08540 (US). ELAN CORPORATION, PLC [IE/IE]; Lincoln House, Lincoln Place, Dublin 2 (IE).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): ALVAREZ, Vernon, L. [US/US]; 187 Rice Drive, Morrisville, PA 19067 (US). O'MAHONY, Daniel, J. [IE/IE]; 75 Avoca Park, Avoca Avenue, Blackrock, Dublin (IE). LAMBKIN, Imelda, J. [IE/IE]; 9 Station Road, Sutton, Dublin 13 (IE). PATERSON, Catherine, A. [GB/IE]; 3 Grange Crescent, Pottery Road, Dunlaoghaire, Dublin (IE). SINGLETON, Ju-			dith [US/US]; Knoll Way, Rocky Way, NJ 08553 (US). BELINKA, Benjamin, A., Jr. [US/US]; 15 Pelham Road, Kendall Park, NJ 08824 (US). CARTER, John, M. [US/US]; 35 Chicory Lane, Trenton, NJ 08638-1926 (US). CAGNEY, Gerard, M. [IE/US]; 2618 Yale Avenue East, Seattle, WA 98102 (US).
(74) Agents: ANTLER, Adriane, M. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).			
(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, GH, GW, HU, ID, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).			
Published			Without international search report and to be republished upon receipt of that report.

(54) Title: RANDOM PEPTIDES THAT BIND TO GASTRO-INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METHODS

(57) Abstract

This invention relates to proteins (e.g., peptides) that are capable of facilitating transport of an active agent through a human or animal gastro-intestinal tissue, and derivatives (e.g., fragments) and analogs thereof, and nucleotide sequences coding for said proteins and derivatives. The proteins of the invention have use in facilitating transport of active agents from the luminal side of the GIT into the systemic blood system, and/or in targeting active agents to the GIT. Thus, for example, by binding (covalently or noncovalently) a protein of the invention to an orally administered drug, the drug can be targeted to specific receptor sites or transport pathways which are known to operate in the human gastro-intestinal tract, thus facilitating its absorption into the systemic system.

20	40	60
MGMKSNSFFGYPLSIIFFIV VNEFCERFSYYGMRAILILY FTNFISWDONLSTAIYHTFV		
80	100	120
ALCYLTPILGALIADSWLGK FKTIVSLSIVYTIGQAVATSV SSINDLTDNHHDGTPDSLKV		
140	160	180
HVVLSLIGLALIALGTGGIK PCVSAFGGDQFEEGQEKQRN RFFSIFYLAINAGSLLSTII		
200	220	240
TPMLRVQOCGIHSKQACYPL AFGVPAALMAVALIVFVLGS GMYKKFKPQGNIMGKVAKCI		
260	280	300
GFAIKNRFRHRSKAFPKREH WLDWAKEKYDERLISQIKMV TRVMFLYIPLPMFWALFDQO		
320	340	360
GSRWTLOATTMSGKIGALEI QPDQMQTVNAILIVIMVPID DAVLYPLIAKCGFNFTSLKK		
380	400	420
MAVGMVLASMAFVVAIIVQV EIDKTLPVFPKGNEVQIKVL NIGNNTMNISLPGEMVTLGP		
440	460	480
MSQTNAFMTFDVNKLTRINI SSPGSPVTAVTDDFKQGQRH TLLVWAPNHYQVVKGQLNQK		
500	520	540
PEKGENGIRFVNTFNLITI TMSGKVYANISSYNASTYQF FPSGIGKFTISSTEIPPPQC		
560	580	600
PNFNTFYLEFGSAYTYIVQR KNDSCPEVKVFEDISANTVN MALQIPQYFLLTGEGVVFV		
620	640	660
TGLEFSYSQAPSNSMKSLOA GWLLTVAVGNIIVLIVAGAG QFSKQWAHEYILFAALLLVC		
680	700	708
V1FA1MAREFYTYINPAE1EA QFEDDFKKNRIEKSNPYFMS GANSOKOM		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

RANDOM PEPTIDES THAT BIND TO GASTRO-INTESTINAL  
TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METHODS

5

This application claims priority to U.S. provisional application Serial No. 60/046,595 filed May 15, 1997, which is incorporated by reference herein in its entirety.

10

1. INTRODUCTION

The present invention relates generally to random peptides capable of specific binding to gastro-intestinal tract (GIT) transport receptors. In particular, this 15 invention relates to peptide sequences and motifs, as well as derivatives thereof, which enhance drug delivery and transport through tissue, such as epithelial cells lining the luminal side of the gastro-intestinal tract (GIT). Production of peptides, derivatives and antibodies is also 20 provided. The invention further relates to pharmaceutical compositions, formulations and related methods.

2. BACKGROUND OF THE INVENTION

2.1. Peptide Libraries

25 There have been two different approaches to the construction of random peptide libraries. According to one approach, peptides have been chemically synthesized *in vitro* in several formats. Examples of chemically synthesized libraries can be found in Fodor, S., et al., 1991, *Science* 251: 767-773; Houghten, R., et al., 1991, *Nature* 354: 84-86; 30 and Lam, K., et al., 1991, *Nature* 354: 82-84.

A second approach to the construction of random peptide libraries has been to use the M13 phage, and, in particular, protein pIII of M13. The viral capsid protein of 35 M13, protein III (pIII), is responsible for infection of bacteria. Several investigators have determined from mutational analysis that the 406 amino acid long pIII capsid

protein has two domains. The C-terminus anchors the protein to the viral coat, while portions of the N-terminus of pIII are essential for interaction with the *E. coli* pillin protein (Crissman, J.W. and Smith, G.P., 1984, *Virology* 132: 445-5 455). Although the N-terminus of the pIII protein has shown to be necessary for viral infection, the extreme N-terminus of the mature protein does tolerate alterations. In 1985, George Smith published experiments reporting the use of the pIII protein of bacteriophage M13 as an experimental system 10 for expressing a heterologous protein on the viral coat surface (Smith, G.P., 1985, *Science* 228: 1315-1317). It was later recognized, independently by two groups, that the M13 phage pIII gene display system could be a useful one for mapping antibody epitopes (De la Cruz, V., et al., 1988, 15 *J. Biol. Chem.* 263: 4318-4322; Parmley, S.F. and Smith, G.P., 1988, *Gene* 73: 305-318).

Parmley, S.F. and Smith, G.P., 1989, *Adv. Exp. Med. Biol.* 251: 215-218 suggested that short, synthetic DNA segments cloned into the pIII gene might represent a library 20 of epitopes. These authors reasoned that since linear epitopes were often ~6 amino acids in length, it should be possible to use a random recombinant DNA library to express all possible hexapeptides to isolate epitopes that bind to antibodies. Scott, J.K. and Smith, G.P., 1990, *Science* 249: 25 386-390 describe construction and expression of an "epitope library" of hexapeptides on the surface of M13. Cwirla, S.E., et al., 1990, *Proc. Natl. Acad. Sci. USA* 87: 6378-6382 also described a somewhat similar library of hexapeptides expressed as gene pIII fusions of M13 fd phage. PCT 30 Application WO 91/19818 published December 26, 1991 by Dower and Cwirla describes a similar library of pentameric to octameric random amino acid sequences. Devlin et al., 1990, *Science*, 249: 404-406, describes a peptide library of about 15 residues generated using an (NNS) coding scheme for 35 oligonucleotide synthesis in which S is G or C. Christian and colleagues have described a phage display library,

expressing decapeptides (Christian, R.B., et al., 1992, J. Mol. Biol. 227: 711-718).

Other investigators have used other viral capsid proteins for expression of non-viral DNA on the surface of 5 phage particles. For example, the major capsid protein pVIII was so used by Cesareni, G., 1992, FEBS Lett. 307: 66-70. Other bacteriophage than M13 have been used to construct peptide libraries. Four and six amino acid sequences corresponding to different segments of the Plasmodium 10 falciparum major surface antigen have been cloned and expressed in the filamentous bacteriophage fd (Greenwood, J., et al., 1991, J. Mol. Biol. 220: 821-827).

Kay et al., 1993, Gene 128: 59-65 (Kay) discloses a method of constructing peptide libraries that encode peptides 15 of totally random sequence that are longer than those of any prior conventional libraries. The libraries disclosed in Kay encode totally synthetic random peptides of greater than about 20 amino acids in length. Such libraries can be advantageously screened to identify peptides, polypeptides 20 and/or other proteins having binding specificity for a variety of ligands. (See also U.S. Patent No. 5,498,538 dated March 12, 1996; and PCT Publication No. WO 94/18318 dated August 18, 1994.)

A comprehensive review of various types of peptide 25 libraries can be found in Gallop et al., 1994, J. Med. Chem. 37:1233-1251.

Screening of peptide libraries has often been done using an antibody as ligand (Parmley and Smith, 1989, Adv. Exp. Med. Biol. 251:215-218; Scott and Smith, 1990, 30 Science 249:386-390). In many cases, the aim of the screening is to identify peptides from the library that mimic the epitopes to which the antibodies are directed. Thus, given an available antibody, peptide libraries are excellent sources for identifying epitopes or epitope-like molecules of 35 that antibody (Yayon et al., 1993, Proc. Natl. Acad. Sci. USA 90:10643-10647).

McCafferty et al., 1990, *Nature* 348:552-554 used PCR to amplify immunoglobulin variable (V) region genes and cloned those genes into phage expression vectors. The authors suggested that phage libraries of V, diversity (D), 5 and joining (J) regions could be screened with antigen. The phage that bound to antigen could then be mutated in the antigen-binding loops of the antibody genes and rescreened. The process could be repeated several times, ultimately giving rise to phage which bind the antigen strongly.

10 Marks et al., 1991, *J. Mol. Biol.* 222:581-597 also used PCR to amplify immunoglobulin variable (V) region genes and cloned those genes into phage expression vectors.

Kang et al., 1991, *Proc. Natl. Acad. Sci. USA* 88:4363-4366 created a phagemid vector that could be used to 15 express the V and constant (C) regions of the heavy and light chains of an antibody specific for an antigen. The heavy and light chain V-C regions were engineered to combine in the periplasm to produce an antibody-like molecule with a functional antigen binding site. Infection of cells 20 harboring this phagemid with helper phage resulted in the incorporation of the antibody-like molecule on the surface of phage that carried the phagemid DNA. This allowed for identification and enrichment of these phage by screening with the antigen. It was suggested that the enriched phage 25 could be subject to mutation and further rounds of screening, leading to the isolation of antibody-like molecules that were capable of even stronger binding to the antigen.

Hoogenboom et al., 1991, *Nucleic Acids Res.* 19:4133-4137 suggested that naive antibody genes might be 30 cloned into phage display libraries. This would be followed by random mutation of the cloned antibody genes to generate high affinity variants.

Bass et al., 1990, *Proteins: Struct. Func. Genet.* 8:309-314 fused human growth hormone (hGH) to the carboxy 35 terminus of the gene III protein of phage fd. This fusion protein was built into a phagemid vector. When cells carrying the phagemid were infected with a helper phage,

about 10% of the phage particles produced displayed the fusion protein on their surfaces. These phage particles were enriched by screening with hGH receptor-coated beads. It was suggested that this system could be used to develop mutants of hGH with altered receptor binding characteristics.

5 Lowman et al., 1991, *Biochemistry* 30:10832-10838 used an improved version of the system of Bass et al. described above to select for mutant hGH proteins with exceptionally high affinity for the hGH receptor. The 10 authors randomly mutagenized the hGH-pIII fusion proteins at sites near the vicinity of 12 amino acids of hGH that had previously been identified as being important in receptor binding.

Balass et al., 1993, *Proc. Natl. Acad. Sci. USA* 15 90:10638-10642 used a phage display library to isolate linear peptides that mimicked a conformationally dependent epitope of the nicotinic acetylcholine receptor. This was done by screening the library with a monoclonal antibody specific for the conformationally dependent epitope. The monoclonal 20 antibody used was thought to be specific to the acetylcholine receptor's binding site for its natural ligand, acetylcholine.

## 2.2. Drug Delivery Systems

25 The common routes of therapeutic drug administration are oral ingestion or parenteral (intravenous, subcutaneous and intramuscular) routes of administration. Intravenous drug administration suffers from numerous 30 limitations, including (i) the risk of adverse effects resulting from rapid accumulation of high concentrations of drug, (ii) repeated injections which can cause patient discomfort; and (iii) the risk of infection at the site of repeated injections. Subcutaneous injection is not generally suitable for delivering large volumes or for irritating 35 substances. Whereas oral administration is generally more convenient, it is limited where the therapeutic agent is not efficiently absorbed by the gastrointestinal tract. To date,

the development of oral formulations for the effective delivery of peptides, proteins and macromolecules has been an elusive target. Poor membrane permeability, enzymatic instability, large molecular size, and hydrophilic properties 5 are four factors that have remained major hurdles for peptide and protein formulations (reviewed by Fix, J.A., 1996, J. Pharmac. Sci. 85:1282-1285). In order to develop an efficacious oral formulation, the peptide must be protected from the enzymatic environment of the gastrointestinal tract 10 (GIT), presented to the absorptive epithelial barrier in a sufficient concentration to effect transcellular flux (Fix, J.A., 1996, J. Pharmac. Sci. 85:1282-1285), and if possible "smuggled" across the epithelial barrier in an apical to basolateral direction.

15 Site specific drug delivery or drug targeting can be achieved at different levels, including (i) primary targeting to a specific organ, (ii) secondary targeting to a specific cell type within that organ and (iii) tertiary targeting where the drug is delivered to specific 20 intracellular structures (e.g., the nucleus for genes) (reviewed in Davis and Jllum, 1994, In: Targeting of Drugs 4, (Eds), Gregoriadis, McCormack and Poste, 183-194). At present there is a considerable amount of ongoing research work in the Drug Delivery Systems (DDS) area, and much of it 25 addresses (i) targeting delivery and (ii) the development of non-invasive ways of getting macromolecules, peptides, proteins, products of the biotechnology industry, etc. into the body (Evers, P., 1995, Developments in Drug Delivery: Technology and Markets, Financial Times Management Report). 30 It is generally accepted that targeted drug delivery is crucial to the improved treatment of certain diseases, especially cancer, and not surprisingly many of the approaches to targeted drug delivery are focused in the cancer area. Many anticancer drugs are toxic to the body as 35 well as to malignant cells. If a drug, or a delivery system, can be modified so that it "homes in" on the tumor, then by maximizing the drug concentration at the disease site, the

anti-cancer effect can be exploited to the full, while toxicity is greatly reduced. Tumors contain antigens which provoke the body to respond by producing antibodies designed to attach to the antigens and destroy them. Monoclonal 5 antibodies are being used as both delivery vehicles targeted to tumor cells (reviewed by Pietersz, G.A., 1990, Bioconjugate Chem. 1:89-95) and as imaging agents to carry molecules of drug or imaging agent to the tumor surface.

10        2.3. Transport Pathways

The epithelial cells lining the luminal side of the GIT are a major barrier to drug delivery following oral administration. However, there are four recognized transport pathways which can be exploited to facilitate drug delivery 15 and transport: the transcellular, paracellular, carrier-mediated, and transcytotic pathways. The ability of a conventional drug, peptide, protein, macromolecule or nano- or microparticulate system to "interact" with one of these transport pathways may result in increased delivery of that 20 drug or particle from the GIT to the underlying circulation.

In the case of the receptor-mediated, carrier-mediated or transcytotic transport pathways, some of the uptake signals have been identified. These signals include, *inter alia*, folic acid, which interacts with the folate 25 receptor, and cobalamin, which interacts with Intrinsic Factor. In addition, leucine- and tyrosine-based peptide sorting motifs or internalization sequences exist, such as YSKV, FPHL, YRGV, YQTI, TEQF, TEVM, TSAF, and YTRF (SEQ ID NOS:203, 204, 205, 206, 207, 208, 209, and 210, 30 respectively), which facilitate uptake or targeting of proteins using specific membrane receptors or binding sites to identify peptides that bind specifically to the receptor or binding site.

Non-receptor based assays to discover particular 35 ligands have also been used. For instance, a strategy for identifying peptides that alter cellular function by scanning whole cells with phage display libraries is disclosed in Fong

et al., Drug Development Research 33:64-70 (1994). However, because whole cells, rather than intact tissue or polarized cell cultures, are used for screening phage display libraries, this procedure does not provide information 5 regarding sequences whose primary function includes affecting transport across polarized cell layers.

Additionally, Stevenson et al., Pharmaceutical Res. 12(9), S94 (1995) discloses the use of Caco-2 monolayers to screen a synthetic tripeptide combinatorial library for 10 information relating to the permeability of di- and tri-peptides.

A method of identifying a peptide which permits or facilitates the transport of an active agent through human or animal tissues has been developed (see U.S. patent 15 application Serial No. 08/746,411 filed November 8, 1996, which is incorporated by reference herein in its entirety). Phage from a random phage library is plated onto or brought into contact with a first side, preferably the apical side, of a tissue sample, either *in vitro*, *in vivo* or *in situ*, or 20 polarized tissue cell culture. The phage which is transported to a second side of the tissue opposite the first side, preferably the basolateral side, is harvested to select transported phages. The transported phages are amplified in a host and this cycle is repeated (using the transported 25 phage from the most recent cycle) to obtain a selected phage library containing phage which can be transported from the first side to the second side.

Discussion or citation of a reference hereinabove shall not be construed as meaning that such reference is 30 prior art to the present invention.

### 3. SUMMARY OF THE INVENTION

The present invention relates generally to random peptides and peptide motifs capable of specific binding to 35 GIT transport receptors. Such proteins can be identified using any random peptide library, e.g., a chemically synthesized peptide library or a biologically expressed

peptide library. If a biological peptide expression library is used, the nucleic acid which encodes the peptide which binds to the ligand of choice can be recovered, and then sequenced to determine its nucleotide sequence and hence deduce the amino acid sequence that mediates binding.

Alternatively, the amino acid sequence of an appropriate binding domain can be determined by direct determination of the amino acid sequence of a peptide selected from a peptide library containing chemically synthesized peptides. In a less preferred aspect, direct amino acid sequencing of a binding peptide selected from a biological peptide expression library can also be performed.

In particular, this invention relates to proteins (e.g., peptides) that are capable of facilitating transport of an active agent through a human or animal gastrointestinal tissue, and derivatives (e.g., fragments) and analogs thereof, and nucleotide sequences coding for said proteins and derivatives.

Preferably, the tissue through which transport is facilitated is of the duodenum, jejunum, ileum, ascending colon, transverse colon, descending colon, or pelvic colon. The tissue is most preferably epithelial cells lining the luminal side of the GIT.

The proteins of the invention have use in facilitating transport of active agents from the luminal side of the GIT into the systemic blood system, and/or in targeting active agents to the GIT. Thus, for example, by binding (covalently or noncovalently) a protein of the invention to an orally administered drug, the drug can be targeted to specific receptor sites or transport pathways which are known to operate in the human gastrointestinal tract, thus facilitating its absorption into the systemic system.

The invention also relates to derivatives and analogs of the invention which are functionally active, i.e., they are capable of displaying one or more known functional activities associated with a full-length peptide. Such

functional activities include but are not limited to antigenicity (ability to bind or to compete with GIT transport receptor-binding peptides for binding to an anti-GIT transport receptor antibody) and ability to bind or 5 compete with full-length peptide for binding to a GIT transport receptor.

The invention further relates to fragments of (and derivatives and analogs thereof) GIT transport receptor-binding peptides which comprise one or more motifs of a GIT 10 transport receptor-binding peptide.

Antibodies to GIT transport receptor-binding peptides and GIT transport receptor-binding peptide derivatives and analogs are additionally provided.

Methods of production of the GIT transport 15 receptor-binding peptides, derivatives, fragments and analogs, e.g., by recombinant means, are also provided.

The present invention also relates to therapeutic methods, pharmaceutical compositions and formulations based on GIT transport receptor-binding peptides. Formulations of 20 the invention include but are not limited to GIT transport receptor-binding peptides or motifs and derivatives (including fragments) thereof; antibodies thereto; and nucleic acids encoding the GIT transport receptor-binding peptides or derivatives associated with an active agent. 25 Preferably, the active agent is a drug or drug-containing nano- or microparticle.

The GIT transport-receptor binding proteins of the invention can also be used to determine levels of the GIT transport receptors in a sample by binding thereto.

30 The GIT transport-receptor binding proteins can also be used to identify molecules that bind thereto, by contacting candidate test molecules under conditions conducive to binding, and detecting any binding that occurs.

#### 35 4. DESCRIPTION OF THE FIGURES

**Figure 1.** Figure 1 shows the human PEPT1 predicted amino acid sequence determined from the sequence of the cDNA clone

coding for human PEPT1 (SEQ ID NO:176) (Liang R. et al. J. Biol. Chem. 270(12):6456-6463 (1995)), including the extracellular domain from amino acid 391 to 573 (Fei et al., Nature 368:563 (1994)).

5 **Figures 2A-2C.** Figures 2A-2C show the DNA sequence of the cDNA coding for the human intestinal peptide-associated transporter HPT1 and the corresponding putative amino acid sequence (bases 1 to 3345; Medline:94204643) (SEQ ID NOS: 177 and 178, respectively).

10 **Figures 3A-3B.** Figures 3A-3B show the putative Human Sucrase-isomaltase complex(hSI) amino acid sequence determined from the sequence of the cDNA clone coding for human sucrase-isomaltase complex (SEQ ID NO:179) (Chantret I., et al., Biochem. J. 285(Pt 3):915-923 (1992)).

15 **Figures 4A-4B.** Figures 4A-4B show the D2H nucleotide and deduced amino acid sequence for the human D2H transporter (SEQ ID NOS:180 and 181, respectively) (Wells, R.G. et al., J. Clin. Invest. 90:1959-1963 (1993)).

20 **Figures 5A-5C.** Figure 5A is a schematic summary of the cloning of the DNA insert present in gene III of the phages selected from the phage display libraries into the expression vector pGex-4T-2. The gene insert in gene III of the phages was amplified by PCR using DNA primers which flank the gene insert and which contained recognition sequences for specific 25 restriction endonucleases at their extreme 5' sides.

Alternatively, specific primers which amplify specific regions of the DNA inserts in gene III of the phages, and which contained recognition sequences for specific restriction endonucleases at their extreme 5' sides, were 30 used in PCR amplification experiments. Following amplification of the gene inserts, the amplified PCR fragments were digested with the restriction endonucleases Xho1 and Not1. Similarly the plasmid pGex-4T-2, which codes for the reporter protein glutathione S-transferase (GST), was 35 digested with the restriction endonucleases Sal1 and Not1. The digested PCR fragments were ligated into the digested plasmid pGex-4T-2 using T4 DNA Ligase and the ligated

products were transformed into competent *Escherichia coli*, with selection of transformants on agar plates containing selection antibiotic. The selected clones were cultured, the plasmids were recovered and the in-frame sequence of the DNA 5 insert in the plasmids was confirmed by DNA sequencing. The correct clones were subsequently used for expression of the GST-fusion proteins (SEQ ID NO:182); Figure 5B shows the series of full-length P31 (designated P31) (SEQ ID NO:43) and truncated peptides derived from P31 (clones # 101, 102, 103 10 and 119), (SEQ ID NOS:183, 184, 185, and 186, respectively) full-length PAX2 (designated PAX2) (SEQ ID NO:55) and truncated peptides derived from PAX2 (clones # 104, 105, 106) (SEQ ID NOS:170, 187, and 188, respectively) and full-length DCX8 (DCX8) (SEQ ID NO:23) and series of truncated peptides 15 derived from DCX8 (clones # 107, 108, 109) (SEQ ID NOS:189, 190, and 191, respectively) that were expressed as fusion proteins to GST. The construction of these GST-fusion proteins is shown in Figure 5A. Figure 5C shows the series of full-length P31 (designated P31) (SEQ ID NO:43) and 20 truncated peptides derived from P31 (clones # 103, 110, 119, 111, and 112) (SEQ ID NOS:185, 192, 193, 194, and 195, respectively), full-length PAX2 (designated PAX2) (SEQ ID NO:55) and truncated peptides derived from PAX2 (clones # 106, 113, 114, 115) (SEQ ID NOS:188, 196, 197, and 198, 25 respectively) and full-length SNi10 (designated SNi10) (SEQ ID NO:4) and series of truncated peptides derived from SNi10 (clones # 116, 117, 118) (SEQ ID NOS:199, 200, and 201, respectively) that were expressed as fusion proteins to GST. The construction of these GST-fusion proteins is shown in 30 Figure 5A. (Underlining and bold in Figs. 5A-5C are for orientation of the sequences.)

**Figures 6A-6B.** Figures 6A-6B show the binding of GST and GST-fusion proteins to recombinant hSI and to fixed C2BBe1 fixed cells as detected by ELISA assays. Figure 6A shows the 35 binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from SNi10 (designated GST-SNi10) and SNi34 (designated GST-SNi34) to

recombinant hSI. Figure 6B shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from SNi10 (designated GST-SNi10) and SNi34 (designated GST-SNi34) to fixed C2BBe1 cells.

5 **Figures 7A-7M.** Figures 7A-7M show the binding of GST peptide and truncated fusion proteins to fixed Caco-2 cells, fixed C2BBe1 cells, and fixed A431 cells or to recombinant GIT transport receptors D2H, HPT1, hPEPT1 or to BSA using increasing concentrations (expressed as  $\mu$ g/ml on the X-axis) 10 of the control GST protein and the GST-fusion proteins, as detected by ELISA assays. Figure 7A shows the binding of the control protein GST, which does not contain a fusion peptide, and the series of GST-fusion proteins from P31 including the fusion to full-length P31 peptide (designated P31) (SEQ ID 15 NO:43) and clone # 101 (designated P31,101), clone # 102 (designated P31, 102) and clone # 103 (designated P31,103). Figure 7B shows the binding of the control protein GST, which does not contain a fusion peptide, and the series of GST-fusion proteins from PAX2 including the fusion to full-length 20 PAX2 peptide (designated PAX2) and clone # 104 (designated PAX2,104), clone # 105 (designated PAX2, 105) and clone # 106 (designated PAX2,106) (SEQ ID NOS:55, 170, 187, and 188, respectively). Figure 7C shows the binding of the control protein GST, which does not contain a fusion peptide, and the 25 series of GST-fusion proteins from DCX8 including the fusion to full-length DCX8 peptide (designated DCX8) and clone # 107 (designated DCX8,107), clone # 108 (designated DCX8, 108) and clone # 109 (designated DCX8,109) (SEQ ID NOS:23, 189, 190, and 191, respectively). Figure 7D shows the binding of the 30 control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from DCX8 (designated GST-DCX8) and DCX11 (designated GST-DCX11) to recombinant D2H. Figure 7E shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins 35 from DCX8 (designated GST-DCX8) and DCX11 (designated GST-DCX11) to fixed C2BBe1 cells. Figure 7F shows the binding of the control protein GST, which does not contain a fusion

peptide, and the GST-fusion proteins from P31 (designated GST-P31) and 5PAX5 (designated GST-5PAX5) to recombinant hPEPT1. Figure 7G shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-5 fusion proteins from P31 (designated GST-P31) and 5PAX5 (designated GST-5PAX5) to fixed C2BBe1 cells. Figure 7H shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from HAX42 (designated GST-HAX42) and PAX2 (designated GST-PAX2) to recombinant HPT1. Figure 7I shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from HAX42 (designated GST-HAX42) and PAX2 (designated GST-PAX2) to fixed C2BBe1 cells. Figure 7J shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from P31 (designated GST-P31) and truncated derivatives clone # 101 (designated GST-P31-101), clone # 102 (designated GST-P31-102), clone # 103 (designated GST-P31-103) to either recombinant hPEPT1 or to BSA. Figure 7K shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from P31 (designated GST-P31) and truncated derivatives clone # 101 (designated GST-P31-101), clone # 102 (designated GST-P31-102), clone # 103 (designated GST-P31-103) to either fixed C2BBe1 cells or to fixed A431 cells. Figure 7L shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from PAX2 (designated GST-PAX2) and truncated derivatives clone # 104 (designated GST-PAX2-104), clone # 105 (designated GST-PAX2-105), clone # 106 (designated GST-PAX2-106) to either recombinant hPEPT1 or to BSA. Figure 7M shows the binding of the control protein GST, which does not contain a fusion peptide, and the GST-fusion proteins from PAX2 (designated GST-PAX2) and truncated derivatives clone # 106 (designated GST-PAX2-106) to either fixed Caco-2 cells or to fixed A431 cells.

**Figures 8A-8D.** Figure 8 shows the transport of GST or GST-peptide fusion derivatives across polarized Caco-2 cells in

an apical to basolateral direction as a function of time (1-4 hours) as detected by ELISA assays. Figure 8A shows the transport of either GST, the GST fusion to full-length P31 peptide (designated P31) (SEQ ID NO:43) and the GST clone 5 derivative clone # 103 (designated P31.103) across polarized Caco-2 cells in an apical to basolateral as a function of time (in hours) following initial administration of the proteins to the apical medium of polarized Caco-2 cells. The line designated No Protein corresponds to control assays in 10 which buffer control was applied to the apical medium of polarized Caco-2 cells followed by sampling of the basolateral medium as a function of time (hours) and assay for GST by the ELISA assay. Figure 8B shows the transport of either GST, the GST fusion to full-length PAX2 peptide 15 (designated PAX2) and the GST clone derivative clone # 106 (designated PAX2.106) across polarized Caco-2 cells in an apical to basolateral as a function of time (in hours) following initial administration of the proteins to the apical medium of polarized Caco-2 cells. The line designated 20 No Protein corresponds to control assays in which buffer control was applied to the apical medium of polarized Caco-2 cells followed by sampling of the basolateral medium as a function of time (hours) and assay for GST by the ELISA assay. Figure 8C shows the transport of either GST, the GST 25 fusion to full-length DCX8 peptide (designated DCX8), and the GST clone derivatives clone # 107 (designated DCX8.107) and clone # 109 (designated DCX8.109) across polarized Caco-2 cells in an apical to basolateral as a function of time (in hours) following initial administration of the proteins to 30 the apical medium of polarized Caco-2 cells. The line designated No Protein corresponds to control assays in which buffer control was applied to the apical medium of polarized Caco-2 cells followed by sampling of the basolateral medium as a function of time (hours) and assay for GST by the ELISA 35 assay. Figure 8D shows the amount of the GST and GST-fusion proteins (GST fusions to P31, P31-103, PAX2, PAX2.106, DCX8, DCX8-107, DCX8-109), used in the experiments shown in panels

A-C above, in the apical medium of the polarized Caco-2 cells as detected by ELISA assay.

**Figures 9A-9B.** Figures 9A-9B show the inhibition of GST-P31 binding to C2BBe1 fixed cells with varying concentration of 5 competitors while holding the concentration of GST-P31 constant at 0.015  $\mu$ M; the peptide competitors are ZElan024 which is the dansylated peptide version of P31 (SEQ ID NO:43) and ZElan044, ZElan049 and ZElan050 which are truncated, dansylated pieces of P31 (SEQ ID NO:43). Data is presented 10 as O.D. versus peptide concentration (Figure 9A) and as percent inhibition of GST-P31 binding versus peptide concentration (Figure 9B).

**Figures 10A-10C.** Figures 10A-10C present a compilation of the results of competition ELISA studies of GST-P31, GST- 15 PAX2, GST-SNi10 and GST-HAX42 versus listed dansylated peptides on fixed C2BBe1 cells ("Z" denotes  $\epsilon$ -amino dansyl lysine). The pI of the dansylated peptides is also included. Estimated  $IC_{50}$  values are in  $\mu$ M and where present,  $IC_{50}$  ranges refer to results from multiple assays. If the  $IC_{50}$  value 20 could not be determined, a ">" or "<" symbol is used. The GST/C2BBe1 column shows GST protein binding to fixed C2BBe1 cells.

**Figures 11A-11B.** Figure 11A shows the transport of GST or GST-peptide fusion derivatives across polarized Caco-2 cells 25 in an apical to basolateral direction at 0, 0.5, 2 and 4 hours as detected by ELISA assays and described elsewhere in the text in full detail. The proteins used in the assay included GST, GST-P31 fusion, GST-5PAX5 fusion, GST-DCX8 fusion, GST-DCX11 fusion, GST-PAX2 fusion, GST-HAX42 fusion, 30 GST-SNi34 fusion and GST-SNi10 fusion. The column designated No protein refers to control experiments in which buffer was applied to the apical medium of the cells and ELISA assay was performed on the corresponding basolateral medium of these cells at 0, 0.5, 2 and 4 hours post buffer addition. Figure 35 11B shows the internalization of GST or GST-peptide fusion derivatives within polarized Caco-2 cells following administration of the GST or GST-fusion protein derivatives

to the apical medium of polarized Caco-2 cells and subsequent recovery of the cells from the transwells and detection of the GST or GST fusions within the recovered cell lysates as detected by ELISA assays and as described elsewhere in the 5 text in full detail. The proteins used in the assay included GST, GST-P31 fusion, GST-5PAX5 fusion, GST-DCX8 fusion, GST-DCX11 fusion, GST-PAX2 fusion, GST-HAX42 fusion, GST-SNi34 fusion and GST-SNi10 fusion. The column designated No protein refers to control experiments in which buffer was 10 applied to the apical medium of the cells and ELISA assay was performed on the corresponding cell lysates of these cells at the end of the experiment.

**Figure 12.** Figure 12 shows the binding of GST and GST-fusion proteins to fixed Caco-2 cells, and the corresponding 15 proteins following digestion with the protease Thrombin which cleaves at a recognition site between the GST portion and the fused peptide portion of the GST-fusion protein. The symbol "--" refers to proteins which were not digested with thrombin and the symbol "+" refers to proteins which were digested 20 with thrombin prior to use in the binding assay. The binding of the proteins to the fixed Caco-2 cells was detected by ELISA assays.

**Figures 13A-13B.** Figures 13A-13B show binding of peptide-coated nanoparticles to fixed Caco-2 cells.

**25 Figures 14A-14B.** Figures 14A-14B show the binding of (A) dansylated peptide SNi10 to the purified hSI receptor and BSA and (B) dansylated peptides and peptide-loaded insulin-containing PLGA particles to fixed C2BBel cells. Figure 14B depicts binding of dansylated peptides corresponding to P31 30 (SEQ ID NO:43), PAX2, HAX42, and SNi10 to fixed C2BBel cells, as well as the insulin-containing PLGA particles adsorbed with each of these peptides. Data is presented with background subtracted.

**Figures 15A-15B.** Figure 15 shows the binding of peptide-35 coated particles to A) S100 and B) P100 fractions harvested from Caco-2 cells. The dilution series 1:2 - 1:64 represents particle concentrations in the range 0.0325-0.5  $\mu$ g/well.

Data is presented with background subtracted. The particles are identified as follows: 939, no peptide; 1635, scrambled PAX2; 1726, P31 D-Arg 16-mer (ZElan053); 1756, HAX42; 1757, PAX2; 1758, HAX42/PAX2.

5 **Figures 16A-16B.** Figure 16 shows the binding of dansylated peptides to P100 fractions harvested from Caco-2 cells. Peptides were assayed in the range 0.0032-2.5 µg/well. Data is presented with background subtracted. A) HAX42, P31 D-form (ZElan 053) and scrambled PAX2; B) PAX2, HAX42 and 10 scrambled PAX2.

Figures 17A-17B. Figures 17A and 17B show (A) the systemic blood glucose and (B)insulin levels following intestinal administration of control (PBS); insulin solution; insulin particles; all 8 peptides mix particles and study group 15 peptide-particles according to this invention (100iu insulin loading).

Figures 18A-18B. Figures 18A and 18B show the (A) systemic blood glucose and (B)insulin levels following intestinal administration of control (PBS); insulin solution; insulin 20 particles and study group peptide-particles according to this invention (300iu insulin loading).

**Figure 19.** Figure 19 shows the enhanced plasma levels of leuprolide upon administration of P31 (SEQ ID NO:43) and PAX2 coated nanoparticles loaded with leuprolide relative to 25 subcutaneous injection. Group 1 was administered leuprolide acetate (12.5 µg) subcutaneously. Group 2 was administered intraduodenally uncoated leuprolide acetate particles (600 µg, 1.5 ml). Group 3 was intraduodenally administered leuprolide acetate particles coated with PAX2 (600 µg; 1.5 30 ml). Group 4 was administered intraduodenally leuprolide acetate particles coated with P31 (SEQ ID NO:43) (600 µg, 1.5 ml).

**Figure 20.** Figure 20 lists P31 (SEQ ID NO:43) known protein homologies.

35 **Figures 21A-21C.** Figures 21A-21C list DCX8 known protein homologies.

**Figure 22.** Figure 22 lists DAB10 known protein homologies.

**Figure 23.** Figure 23 shows the DNA sequence (SEQ ID NO:211) and the corresponding amino acid sequence (SEQ ID NO:212) for glutathione S-transferase (Smith and Johnson, 1988, Gene 7:31-40).

5

#### **5. DETAILED DESCRIPTION OF THE INVENTION**

The present invention relates to proteins (e.g., peptides) that bind to GIT transport receptors and nucleic acids that encode such proteins. The invention further 10 relates to fragments and other derivatives of such proteins. Nucleic acids encoding such fragments or derivatives are also within the scope of the invention. The invention further relates to fragments (and derivatives and analogs thereof) of 15 GIT transport receptor-binding peptides which comprise one or more domains of the GIT transport receptor-binding peptides.

The invention also relates to derivatives of GIT transport receptor-binding proteins and analogs of the invention which are functionally active, i.e., they are capable of displaying one or more known functional activities 20 associated with a full-length GIT transport receptor-binding peptide. Such functional activities include but are not limited to ability to bind to a GIT transport receptor, antigenicity [ability to bind (or compete with peptides for binding) to an anti-GIT transport receptor-binding peptide 25 antibody], immunogenicity (ability to generate antibody which binds to GIT transport receptor-binding peptide), etc.

Production of the foregoing proteins and derivatives, by, e.g., recombinant methods, is also provided.

Antibodies to GIT transport receptor-binding 30 proteins, derivatives and analogs, are additionally provided.

The present invention also relates to therapeutic and diagnostic methods and compositions based on GIT transport receptor-binding proteins and nucleic acids.

The invention is illustrated by way of examples 35 *infra*.

For clarity of disclosure, and not by way of limitation, the detailed description of the invention is divided into the subsections which follow.

5        5.1. GIT Transport Receptor-Binding Peptides, Derivatives and Analogs

The invention relates to peptides that bind GIT transport receptors and derivatives (including but not limited to fragments) and analogs thereof. In specific 10 embodiments, of the present invention, such peptides that bind to GIT transport receptor include but are not limited to those containing as primary amino acid sequences, all or part of the amino acid sequences substantially as depicted in Table 7 (SEQ ID NOS:1-55). Nucleic acids encoding such 15 peptides, derivatives and peptide analogs are also provided. In one embodiment, the GIT transport receptor-binding peptides are encoded by the nucleic acids having the nucleotide sequences set forth in Table 8 *infra* (SEQ ID NOS:56-109). Proteins whose amino acid sequence comprise, or 20 alternatively, consist of SEQ ID NOS:1-55 or a portion thereof that mediates binding to a GIT transport receptor are provided.

The production and use of derivatives and analogs related to GIT transport receptor-binding peptides are within 25 the scope of the present invention. In a specific embodiment, the derivative or analog is functionally active, i.e., capable of exhibiting one or more functional activities associated with a full-length GIT transport receptor-binding peptide. For example, such derivatives or analogs which have 30 the desired immunogenicity or antigenicity can be used, in immunoassays, for immunization, etc. A specific embodiment relates to a GIT transport receptor-binding peptide fragment that can be bound by an anti-GIT transport receptor-binding peptide antibody. In a preferred aspect, the derivatives or 35 analogs have the ability to bind to a GIT transport receptor. Derivatives or analogs of GIT transport receptor-binding peptides can be tested for the desired activity by procedures

known in the art, including binding to a GIT transport receptor domain or to Caco-2 cells, *in vitro*, or to intestinal tissue, *in vivo*. (See the Examples *infra*.)

In particular, derivatives can be made by altering 5 GIT transport receptor-binding peptide sequences by substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other nucleotide sequences which encode substantially the same amino acid sequence may be used 10 in the practice of the present invention. These include but are not limited to nucleotide sequences which are altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the GIT 15 transport receptor-binding peptide derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of a GIT transport receptor-binding peptide including altered sequences in which functionally equivalent 20 amino acid residues are substituted for residues within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent 25 alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and 30 methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and 35 glutamic acid.

In a specific embodiment of the invention, proteins consisting of or, alternatively, comprising all or a fragment

of a GIT transport receptor-binding peptide consisting of at least 5, 10, 15, 20, 25, 30 or 35 (contiguous) amino acids of the full-length GIT transport receptor-binding peptide are provided. In a specific embodiment, such proteins are not 5 more than 20, 30, 40, 50, or 75 amino acids in length. Derivatives or analogs of GIT transport receptor-binding peptides include but are not limited to those molecules comprising regions that are substantially homologous to GIT transport receptor-binding peptides or fragments thereof 10 (e.g., at least 50%, 60%, 70%, 80% or 90% identity) (e.g., over an identical size sequence or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art) or whose encoding nucleic acid is capable of hybridizing to a coding GIT transport 15 receptor-binding peptide sequence, under stringent, moderately stringent, or nonstringent conditions.

In a specific embodiment, the GIT transport receptor-binding derivatives of the invention are not known proteins with homology to the GIT transport receptor-binding 20 peptides of the invention or portions thereof.

The GIT transport receptor-binding peptide derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein 25 level. For example, the cloned GIT transport receptor-binding peptide gene sequence can be modified by any of numerous strategies known in the art (Maniatis, T., 1990, Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The 30 sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*. In the production of the gene encoding a derivative or analog of GIT transport receptor-binding peptides, care should be taken to 35 ensure that the modified gene remains within the same translational reading frame uninterrupted by translational

stop signals, in the gene region where the desired GIT transport receptor-binding peptides activity is encoded.

Additionally, nucleic acid sequences encoding the GIT transport receptor-binding peptides can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, including but not limited to, chemical mutagenesis, *in vitro* site-directed mutagenesis (Hutchinson, C., et al., 1978, *J. Biol. Chem.* 253:6551), use of TAB® linkers (Pharmacia), use of PCR primers containing mutation(s) for use in amplification, etc.

Manipulations of GIT transport receptor-binding peptide sequences may also be made at the protein level. Included within the scope of the invention are GIT transport receptor-binding peptide fragments or other derivatives or analogs which are differentially modified during or after translation or chemical synthesis, e.g., by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH<sub>4</sub>; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; etc. In a specific embodiment, the amino- and/or carboxy-termini are modified.

In addition, GIT transport receptor-binding peptides and analogs and derivatives thereof can be chemically synthesized. For example, a peptide corresponding to all or a portion of a GIT transport receptor-binding peptide which comprises the desired domain or which mediates the desired activity *in vitro*, can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical

amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the GIT transport receptor-binding peptide sequence. Non-classical amino acids include but are not limited to the D-isomers of the common 5 amino acids,  $\alpha$ -amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid,  $\gamma$ -Abu,  $\epsilon$ -Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, 10 phenylglycine, cyclohexylalanine,  $\beta$ -alanine, fluoro-amino acids, designer amino acids such as  $\beta$ -methyl amino acids,  $\alpha$ -methyl amino acids,  $\text{N}\alpha$ -methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

15 In a specific embodiment, the GIT transport receptor-binding peptide derivative is a chimeric, or fusion, peptide comprising a GIT transport receptor-binding peptide or fragment thereof (preferably consisting of at least a domain or motif of the GIT transport receptor-binding 20 peptide, or at least 6, 10, 15, 20, 25, 30 or all amino acids of the GIT transport receptor-binding peptides or a binding portion thereof) joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different peptide. In one embodiment, such a chimeric peptide is 25 produced by recombinant expression of a nucleic acid encoding the protein (comprising a transport receptor-coding sequence joined in-frame to a coding sequence for a different protein). Such a chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired 30 amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, e.g., by use of a peptide synthesizer. Chimeric 35 genes comprising portions of GIT transport receptor fused to any heterologous protein-encoding sequences may be constructed. A specific embodiment relates to a chimeric

protein comprising a fragment of GIT transport receptor-binding peptides of at least six amino acids.

In another specific embodiment, the GIT transport receptor-binding peptide derivative is a molecule comprising 5 a region of homology with a GIT transport receptor-binding peptide. By way of example, in various embodiments, a first protein region can be considered "homologous" to a second protein region when the amino acid sequence of the first region is at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, or 10 95% identical, when compared to any sequence in the second region of an equal number of amino acids as the number contained in the first region or when compared to an aligned sequence of the second region that has been aligned by a computer homology program known in the art. For example, a 15 molecule can comprise one or more regions homologous to a GIT transport receptor-binding peptide domain (see *infra*) or a portion thereof.

The GIT transport receptor-binding proteins and derivatives thereof of the invention can be assayed for 20 binding activity by suitable *in vivo* or *in vitro* assays, e.g., as described in the examples *infra* and/or as will be known to the skilled artisan.

Other specific embodiments of derivatives and analogs are described in the subsection below and examples 25 sections *infra*.

#### **5.2. Motifs/Derivatives of GIT Transport Receptor-Binding Peptides Containing One or More Domains of The Protein**

In a specific embodiment, the invention relates to 30 GIT transport receptor-binding peptide derivatives and analogs, in particular GIT transport receptor-binding peptide fragments and derivatives of such fragments, that comprise, or alternatively consist of, one or more domains of a GIT transport receptor-binding peptide. In particular, examples 35 of such domains are identified in the examples *infra*.

### 5.3. Synthesis of Peptides

The peptides and derivatives of the present invention may be chemically synthesized or synthesized using recombinant DNA techniques.

5

#### 5.3.1. Procedure For Solid Phase Synthesis

Peptides may be prepared chemically by methods that are known in the art. For example, in brief, solid phase peptide synthesis consists of coupling the carboxyl group of 10 the C-terminal amino acid to a resin and successively adding N-alpha protected amino acids. The protecting groups may be any known in the art. Before each new amino acid is added to the growing chain, the protecting group of the previous amino acid added to the chain is removed. The coupling of amino acids to appropriate resins is described by Rivier et al., U.S. Patent No. 4,244,946. Such solid phase syntheses have been described, for example, by Merrifield, 1964, J. Am. Chem. Soc. 85:2149; Vale et al., 1981, Science 213:1394-1397; Marki et al., 1981, J. Am. Chem. Soc. 103:3178 and in U.S. 20 Patent Nos. 4,305,872 and 4,316,891. In a preferred aspect, an automated peptide synthesizer is employed.

By way of example but not limitation, peptides can be synthesized on an Applied Biosystems Inc. ("ABI") model 431A automated peptide synthesizer using the "Fastmoc" synthesis protocol supplied by ABI, which uses 2-(1H-Benzotriazol-1-yl)-1,1,3,3,-tetramethyluronium hexafluorophosphate ("HBTU") (R. Knorr et al., 1989, Tet. Lett., 30:1927) as coupling agent. Syntheses can be carried out on 0.25 mmol of commercially available 30 4-(2',4'-dimethoxyphenyl-(9-fluorenyl-methoxycarbonyl)-aminomethyl)-phenoxy polystyrene resin ("Rink resin" from Advanced ChemTech) (H. Rink, 1987, Tet. Lett. 28:3787). Fmoc amino acids (1 mmol) are coupled according to the Fastmoc protocol. The following side chain 35 protected Fmoc amino acid derivatives are used: FmocArg(Pmc)OH; FmocAsn(Mbh)OH; FmocAsp(<sup>t</sup>Bu)OH; FmocCys(Acm)OH; FmocGlu(<sup>t</sup>Bu)OH; FmocGln(Mbh)OH; FmocHis(Tr)OH;

FmocLys(Boc)OH; FmocSer(<sup>t</sup>Bu)OH; FmocThr(<sup>t</sup>Bu)OH; FmocTyr(<sup>t</sup>Bu)OH. [Abbreviations: Acm, acetamidomethyl; Boc, tert-butoxycarbonyl; <sup>t</sup>Bu, tert-butyl; Fmoc, 9-fluorenylmethoxycarbonyl; Mbh, 4,4'-dimethoxybenzhydryl; 5 Pmc, 2,2,5,7,8-pentamethylchroman-6-sulfonyl; Tr, trityl].

Synthesis is carried out using N-methylpyrrolidone (NMP) as solvent, with HBTU dissolved in N,N-dimethylformamide (DMF). Deprotection of the Fmoc group is effected using approximately 20% piperidine in NMP. At 10 the end of each synthesis the amount of peptide present is assayed by ultraviolet spectroscopy. A sample of dry peptide resin (about 3-10 mg) is weighed, then 20% piperidine in DMA (10 ml) is added. After 30 min sonication, the UV (ultraviolet) absorbance of the dibenzofulvene-piperidine 15 adduct (formed by cleavage of the N-terminal Fmoc group) is recorded at 301 nm. Peptide substitution (in mmol g<sup>-1</sup>) can be calculated according to the equation:

$$\text{substitution} = \frac{A \times v}{7800 \times w} \times 1000$$

20 where A is the absorbance at 301 nm, v is the volume of 20% piperidine in DMA (in ml), 7800 is the extinction coefficient (in mol<sup>-1</sup>dm<sup>3</sup>cm<sup>-1</sup>) of the dibenzofulvene-piperidine adduct, and w is the weight of the peptide-resin sample (in mg).

25 Finally, the N-terminal Fmoc group is cleaved using 20% piperidine in DMA, then acetylated using acetic anhydride and pyridine in DMA. The peptide resin is thoroughly washed with DMA, CH<sub>2</sub>Cl<sub>2</sub>, and finally diethyl ether.

### 30 5.3.2. Cleavage And Deprotection

35 By way of example but not limitation, cleavage and deprotection can be carried out as follows: The air-dried peptide resin is treated with ethylmethyl-sulfide (EtSMe), ethanedithiol (EDT), and thioanisole (PhSMe) for approximately 20 min. prior to addition of 95% aqueous trifluoracetic acid (TFA). A total volume of approximately 50 ml of these reagents per gram of peptide-resin is used.

The following ratio is used: TFA:EtSMe:EDT:PhSMe (10:0.5:0.5:0.5). The mixture is stirred for 3 h at room temperature under an atmosphere of N<sub>2</sub>. The mixture is filtered and the resin washed with TFA (2 x 3 ml). The 5 combined filtrate is evaporated in vacuo, and anhydrous diethyl ether added to the yellow/orange residue. The resulting white precipitate is isolated by filtration. See King et al., 1990, Int. J. Peptide Protein Res. 36:255-266 regarding various cleavage methods.

10

#### 5.3.3. Purification of the Peptides

Purification of the synthesized peptides can be carried out by standard methods including chromatography (e.g., ion exchange, affinity, and sizing column 15 chromatography, high performance liquid chromatography (HPLC)), centrifugation, differential solubility, or by any other standard technique.

20

#### 5.3.4. Biological Peptide Libraries

Biological peptide libraries can be used to express and identify peptides that bind to GIT transport receptors. According to this second approach, involving recombinant DNA techniques, peptides can, by way of example, be expressed in biological systems as either soluble fusion proteins or viral 25 capsid proteins.

30

##### 5.3.4.1. Methods To Identify Binders: Construction Of Libraries

In a specific embodiment, the peptides of the invention that specifically bind to GIT transport receptors are identified by screening a random peptide library by contacting the library with a ligand selected from among HPT1, hPEPT1, D2H, or hSI (or a molecule consisting essentially of an extracellular domain thereof or fragment of the domain) to identify members of the library that 35 specifically bind to the ligand.

In a particular embodiment, a process to identify the peptides of the present method utilizes a library of recombinant vectors constructed by methods well known in the art and comprises screening a library of recombinant vectors 5 expressing inserted synthetic oligonucleotide sequences encoding extracellular GIT transport receptor domains, for example, attached to an accessible surface structural protein of a vector to isolate those members producing peptides that bind to HPT1, hPEPT1, D2H, or hSI. The nucleic acid sequence 10 of the inserted synthetic oligonucleotides of the isolated vector is determined and the amino acid sequence encoded can be deduced to identify a binding domain that binds the ligand of choice (e.g., HPT1, hPEPT1, D2H, or hSI).

The present invention encompasses a method for 15 identifying a peptide which binds to a ligand selected from among HPT1, hPEPT1, D2H, or hSI comprising: screening a library of random peptides with the ligand (or an extracellular domain or fragment thereof) under conditions conducive to ligand binding and isolating the peptide which 20 binds to the ligand. Additionally, the methods of the invention further comprise determining the nucleotide sequence encoding the binding domain of the peptide identified to deduce the amino acid sequence of the binding domain.

25

#### 5.3.4.2. Preparation of Extracellular Domain Ligand

In a specific embodiment, molecules consisting essentially of an extracellular domain of the desired GIT 30 transport receptor or a fragment of an extracellular domain are used to screen a random peptide library for binding thereto. Preferably, a nucleic acid encoding the extracellular domain is cloned and recombinantly expressed, and the domain is then purified for use. The GIT transport 35 receptor is preferably selected from among HPT1, hPEPT1, D2H, or hSI.

5.3.4.3. Methods to Identify Binders:  
Screening Libraries

Once a suitable random peptide library has been constructed (or otherwise obtained), the library is screened 5 to identify peptides having binding affinity for the GIT transport receptor, e.g., HPT1, hPEPT1, D2H, or hSI. In a preferred aspect, the library is a TSAR library (see U.S. Patent No. 5,498,538 dated March 12, 1996 and PCT Publication WO 94/18318 dated August 18, 1994, both of which are 10 incorporated by reference herein in their entireties).

Screening the libraries can be accomplished by any of a variety of methods known to those of skill in the art. See, e.g., the following references, which disclose screening of peptide libraries: Parmley and Smith, 1989, *Adv. Exp. Med. Biol.* 251: 215-218; Scott and Smith, 1990, *Science* 249: 386-390; Fowlkes et al., 1992, *BioTechniques* 13: 422-427; Oldenburg et al., 1992, *Proc. Natl. Acad. Sci. USA* 89: 5393-5397; Yu et al., 1994, *Cell* 76: 933-945; Staudt et al., 1988, *Science* 241: 577-580; Bock et al., 1992, *Nature* 355: 564-566; 15 Tuerk et al., 1992, *Proc. Natl. Acad. Sci. USA* 89: 6988-6992; Ellington et al., 1992, *Nature* 355: 850-852; U.S. Patent No. 20 5,096,815, U.S. Patent No. 5,223,409, and U.S. Patent No. 5,198,346, all to Ladner et al.; and Rebar and Pabo, 1993, *Science* 263: 671-673. See also PCT publication WO 94/18318, dated August 18, 1994.

One of ordinary skill in the art will recognize that, with suitable modifications, the screening methods described below would be suitable for a wide variety of biological expression libraries.

Once a library has been constructed or otherwise obtained, the library is screened to identify binding molecules having specific binding affinity for a ligand for a GIT transport receptor preferably selected from among HPT1, hPEPT1, D2H, or hSI.

Screening the libraries can be accomplished by any 35 of a variety of methods known to those of skill in the art. Exemplary screening methods are described in Fowlkes et al.,

1992, *BioTechniques*, 13:422-427 and include contacting the vectors with an immobilized target ligand and harvesting those vectors that bind to said ligand. Such useful screening methods, are designated "panning" methods. In 5 panning methods useful to screen the present libraries, the target ligand can be immobilized on plates, beads (such as magnetic beads), sepharose, beads used in columns, etc. If desired, the immobilized target ligand can be "tagged", e.g., using labels such as biotin, fluorescein isothiocyanate, 10 rhodamine, etc. e.g. for FACS sorting. Panning is also disclosed in Parmley, S.F. and Smith, G.P., 1988, *Gene* 73: 305-318.

In a particular embodiment of the invention, the library can be screened with a recombinant receptor domain.

15 In another embodiment, the library can be screened successively with receptor domains and then on CaCO-2 cells.

For screening of the peptide libraries *in vitro*, the solvent requirements involved in screening are not limited to aqueous solvents; thus, nonphysiological binding 20 interactions and conditions different from those found *in vivo* can be exploited.

Screening a library can be achieved using a method comprising a first "enrichment" step and a second filter lift as follows. The following description is given by way of 25 example, not limitation.

Binders from an expressed library (e.g., in phage) capable of binding to a given ligand ("positives") are initially enriched by one or two cycles of panning or affinity chromatography. A microtiter well is passively 30 coated with the ligand (e.g., about 10 µg in 100 µl). The well is then blocked with a solution of BSA to prevent non-specific adherence of the phage of the library to the plastic surface. For example, about 10<sup>11</sup> phage particles expressing peptides are then added to the well and incubated for several 35 hours. Unbound phage are removed by repeated washing of the plate, and specifically bound phage are eluted using an acidic glycine-HCl solution or other elution buffer. The

eluted phage solution is neutralized with alkali, and amplified, e.g., by infection of *E. coli* and plating on large petri dishes containing Luria broth (LB) in agar. Amplified cultures expressing the binding peptides are then titered and 5 the process repeated. Alternatively, the ligand can be covalently coupled to agarose or acrylamide beads using commercially available activated bead reagents. The phage solution is then simply passed over a small column containing the coupled bead matrix which is then washed extensively and 10 eluted with acid or other eluant. In either case, the goal is to enrich the positives to a frequency of about  $> 1/10^5$ .

Following enrichment, a filter lift assay is conducted. For example, when specific binders are expressed in phage, approximately  $1-2 \times 10^5$  phage are added to 500  $\mu$ l of 15 log phase *E. coli* and plated on a large Luria Broth-agarose plate with 0.7% agarose in broth. The agarose is allowed to solidify, and a nitrocellulose filter (e.g., 0.45  $\mu$ ) is placed on the agarose surface. A series of registration marks is made with a sterile needle to allow re-alignment of 20 the filter and plate following development as described below. Phage plaques are allowed to develop by overnight incubation at 37 °C (the presence of the filter does not inhibit this process). The filter is then removed from the plate with phage from each individual plaque adhered *in situ*. 25 The filter is then exposed to a solution of BSA or other blocking agent for 1-2 hours to prevent non-specific binding of the ligand (or "probe").

The probe itself is labeled, for example, either by biotinylation (using commercial NHS-biotin) or direct enzyme 30 labeling, e.g., with horse radish peroxidase or alkaline phosphatase. Probes labeled in this manner are indefinitely stable and can be re-used several times. The blocked filter is exposed to a solution of probe for several hours to allow the probe to bind *in situ* to any phage on the filter 35 displaying a peptide with significant affinity to the probe. The filter is then washed to remove unbound probe, and then developed by exposure to enzyme substrate solution (in the

case of directly labeled probe) or further exposed to a solution of enzyme-labeled avidin (in the case of biotinylated probe). Positive phage plaques are identified by localized deposition of colored enzymatic cleavage product 5 on the filter which corresponds to plaques on the original plate. The developed filter is simply realigned with the plate using the registration marks, and the "positive" plaques are cored from the agarose to recover the phage. Because of the high density of plaques on the original plate, 10 it may be difficult to isolate a single plaque from the plate on the first pass. Accordingly, phage recovered from the initial core can be re-plated at low density and the process can be repeated to allow isolation of individual plaques and hence single clones of phage.

15         Successful screening experiments are optimally conducted using 3 rounds of serial screening. The recovered cells are then plated at a low density to yield isolated colonies for individual analysis. The individual colonies are selected and used to inoculate LB culture medium 20 containing ampicillin. After overnight culture at 37°C, the cultures are then spun down by centrifugation. Individual cell aliquots are then retested for binding to the target ligand attached to the beads. Binding to other beads having attached thereto a non-relevant ligand, can be used as a 25 negative control.

One aspect of screening the libraries is that of elution. The following discussion is applicable to any system where the random peptide is expressed on a surface fusion molecule. It is conceivable that the conditions that 30 disrupt the peptide-target interactions during recovery of the phage are specific for every given peptide sequence from a plurality of proteins expressed on phage. For example, certain interactions may be disrupted by acid pH but not by basic pH, and *vice versa*. Thus, it may be desirable to test 35 a variety of elution conditions (including but not limited to pH 2-3, pH 12-13, excess target in competition, detergents, mild protein denaturants, urea, varying temperature, light,

presence or absence of metal ions, chelators, etc.) and compare the primary structures of the binding proteins expressed on the phage recovered for each set of conditions to determine the appropriate elution conditions for each 5 ligand/binding protein combination. Some of these elution conditions may be incompatible with phage infection because they are bactericidal and will need to be removed by dialysis (i.e., dialysis bag, Centricon/Amicon microconcentrators).

In a preferred embodiment, a phage display library 10 of random peptides is screened to select phage expressing peptides that bind to a GIT transport receptor. Preferably, a first step is to isolate a preselected phage library. The "preselected phage library" is a library consisting of a subpopulation of a phage display library. This subpopulation 15 can be formed by initially screening against either a target GIT transport receptor (or domain thereof) so as to permit the selection of a subpopulation of phages which specifically bind to the receptor. Alternatively, the subpopulation can be formed by screening against a target cell or cell type or 20 tissue type or tissue barrier of the gastro-intestinal tract, so as to permit the selection of a subpopulation of phages which either bind specifically to the target cell or target cell type or target tissue or target tissue barrier, or which binds to and/or is transported across (or between) the target 25 cell or target cell type or target tissue or target tissue barrier either *in situ* or *in vivo*. This preselected phage library or subpopulation of selected phages can also be rescreened against the target GIT transport receptor, permitting the further selection of a subpopulation of phages 30 which bind to the GIT transport receptor or target cell or target cell type or target tissue or target tissue barrier or which bind to and/or is transported across the target cell, target tissue or target tissue barrier either *in situ* or *in vivo*. Such rescreening can be repeated from zero to 30 times 35 with each successive "pre-selected phage library" generating additional pre-selected phage libraries.

In a preferred embodiment, a preselected phage library binding a ligand that is a GIT transport receptor preferably selected from among HPT1, hPEPT1, D2H, or hSI is obtained by an *in vitro* screening step as described above, 5 and then the phage are optionally further characterized using *in vitro* assays consisting of binding phage directly to the receptor domain of interest or, alternatively, to Caco-2 cells or using *in vivo* assays. In another preferred embodiment, *in vivo* assays are used that measure uptake of 10 phage by intestinal tissue or, alternatively, through the GIT. In alternative embodiments, such further *in vitro* or *in vivo* assays can be used as the initial screening step.

*In vivo* assays that may be used are described in the examples *infra*.

15

**5.4. Generation of Antibodies to GIT Transport Receptor-Binding Peptides and Derivatives Thereof**

According to the invention, a GIT transport receptor-binding peptide, fragments or other derivatives, or 20 analogs thereof, may be used as an immunogen to generate antibodies which immunospecifically bind such an immunogen. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library.

25 Various procedures known in the art may be used for the production of polyclonal antibodies to a GIT transport receptor-binding peptide or derivative or analog. For the production of antibody, various host animals can be immunized by injection with the native GIT transport receptor-binding peptides, or a synthetic version, or derivative (e.g., 30 fragment) thereof, including but not limited to rabbits, mice, rats, fowl, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface 35 active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet

hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *corynebacterium parvum*.

For preparation of monoclonal antibodies directed 5 toward a GIT transport receptor-binding peptide or analog thereof, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (1975, *Nature* 256:495-497), 10 as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, *Immunology Today* 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). In an 15 additional embodiment of the invention, monoclonal antibodies can be produced in germ-free animals utilizing recent technology (PCT/US90/02545). According to the invention, human antibodies may be used and can be obtained by using human hybridomas (Cote et al., 1983, *Proc. Natl. Acad. Sci. U.S.A.* 80:2026-2030) or by transforming human B cells with EBV virus *in vitro* (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, pp. 77-96). According to the invention, techniques developed for the 20 production of "chimeric antibodies" (Morrison et al., 1984, *Proc. Natl. Acad. Sci. U.S.A.* 81:6851-6855; Neuberger et al., 1984, *Nature* 312:604-608; Takeda et al., 1985, *Nature* 314:452-454) by splicing the genes from a mouse antibody molecule specific for GIT transport receptor-binding peptides together with genes from a human antibody molecule of 25 30 appropriate biological activity can be used.

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce GIT transport receptor-binding peptide-specific single chain antibodies. An 35 additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse et al., 1989, *Science* 246:1275-1281) to allow

rapid and easy identification of monoclonal Fab fragments with the desired specificity for GIT transport receptor-binding peptides, derivatives, or analogs.

Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')<sub>2</sub> fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragment, the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent, and Fv fragments.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, e.g. ELISA (enzyme-linked immunosorbent assay). For example, to select antibodies which recognize a specific domain of a GIT transport receptor-binding peptide, one may assay generated hybridomas for a product which binds to a GIT transport receptor-binding peptide fragment containing such a domain.

Antibodies specific to a domain of a GIT transport receptor-binding peptide are also provided.

The foregoing antibodies can be used in methods known in the art relating to the localization and activity of the GIT transport receptor-binding peptide sequences of the invention, e.g., for imaging these peptides after *in vivo* administration (e.g., to monitor treatment efficacy), measuring levels thereof in appropriate physiological samples, in diagnostic methods, etc. For instance, antibodies or antibody fragments specific to a domain of a GIT transport receptor-binding peptide or to a derivative of a peptide, such as a dansyl group or some other epitope introduced into the peptide, can be used to 1) identify the presence of the peptide on a nanoparticle or other substrate; 2) quantify the amount of peptide on the nanoparticle; 3) measure the level of the peptide in appropriate physiological samples; 4) perform immunohistology on tissue

samples; 5) image the peptide after *in vivo* administration; 6) purify the peptide from a mixture using an immunoaffinity column or 7) bind or fix the peptide to the surface of nanoparticle. This last use envisions attaching the antibody 5 (or fragment of the antibody) to the surface of drug-loaded nanoparticles or other substrate and then incubating this conjugate with the peptide. This procedure results in binding of the peptide in a certain fixed orientation, resulting in a particle that contains the peptide bound to 10 the antibody in such a way that the peptide is fully active.

Abtides (or Antigen binding peptides) specific to a domain of a GIT transport receptor-binding peptide or to a derivative of a peptide, such as a dansyl group or some other epitope introduced into the peptide, can be used for the same 15 seven purposes identified above for antibodies.

#### 5.5. Assays of GIT Transport Receptor-Binding Peptides, Derivatives and Analogs

The functional activity of GIT transport receptor- 20 binding peptides, derivatives and analogs can be assayed by various methods.

In a preferred embodiment, in which binding to a GIT transport receptor is being assayed, the binding can be assayed by *in vivo* or *in vitro* assays such as described in 25 the examples *infra*, or by other means that are known in the art.

In another embodiment, where one is assaying for the ability to bind or compete with full-length GIT transport receptor-binding peptide for binding to anti-GIT transport receptor-binding peptide antibody, various immunoassays known 30 in the art can be used, including but not limited to competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, 35 immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (using colloidal gold, enzyme or radioisotope labels, for example), western

blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunolectrophoresis assays, etc. In 5 one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labelled. Many 10 means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention.

Other methods will be known to the skilled artisan and are within the scope of the invention.

15

#### 5.6. Uses

The invention provides compositions comprising the GIT transport receptor-binding proteins of the invention bound to a material comprising an active agent. Such 20 compositions have use in targeting the active agent to the GIT and/or in facilitating transfer through the lumen of the GIT into the systemic circulation. Where the active agent is an imaging agent, such compositions can be administered *in vivo* to image the GIT (or particular transport receptors 25 thereof). Other active agents include but are not limited to: any drug or antigen or any drug- or antigen-loaded or drug- or antigen-encapsulated nanoparticle, microparticle, liposome, or micellar formulation capable of eliciting a biological response in a human or animal. Examples of drug- 30 or antigen-loaded or drug- or antigen-encapsulated formulations include those in which the active agent is encapsulated or loaded into nano- or microparticles, such as biodegradable nano- or microparticles, and which have the GIT transport receptor-binding protein or derivative or analog 35 adsorbed, coated or covalently bound, such as directly linked or linked via a linking moiety, onto the surface of the nano- or microparticle. Additionally, the protein, derivative or

analog can form the nano- or microparticle itself or the protein, derivative or analog can be covalently attached to the polymer or polymers used in the production of the biodegradable nano- or microparticles or drug-loaded or drug-  
5 encapsulated nano- or microparticles or the peptide can be directly conjugated to the active agent. Such conjugations to active agents include fusion proteins in which a DNA sequence coding for the peptide is fused in-frame to the gene or cDNA coding for a therapeutic peptide or protein such that  
10 the modified gene codes for a recombinant fusion protein.

In a preferred embodiment, the invention provides for treatment of various diseases and disorders by administration of a therapeutic compound (termed herein "Therapeutic"). Such "Therapeutics" include but are not  
15 limited to: GIT transport receptor-binding proteins, and analogs and derivatives (including fragments) thereof (e.g., as described hereinabove) that bind to GIT transport receptors, bound to an active agent of value in the treatment or prevention of a disease or disorder (preferably a  
20 mammalian, most preferably human, disease or disorder). Therapeutics also include but are not limited to nucleic acids encoding the GIT transport receptor-binding proteins, analogs, or derivatives bound to such a therapeutic or prophylactic active agent. The active agent is preferably a  
25 drug.

Any drug known in the art may be used, depending upon the disease or disorder to be treated or prevented, and the type of subject to which it is to be administered. As used herein, the term "drug" includes, without limitation,  
30 any pharmaceutically active agent. Representative drugs include, but are not limited to, peptides or proteins, hormones, analgesics, anti-migraine agents, anti-coagulant agents, anti-emetic agents, cardiovascular agents, anti-hypertensive agents, narcotic antagonists, chelating agents,  
35 anti-anginal agents, chemotherapy agents, sedatives, anti-neoplastics, prostaglandins, and antidiuretic agents. Typical drugs include peptides, proteins or hormones such as

insulin, calcitonin, calcitonin gene regulating protein, atrial natriuretic protein, colony stimulating factor, betaseron, erythropoietin (EPO), interferons such as  $\alpha$ ,  $\beta$  or  $\gamma$  interferon, somatropin, somatotropin, somatostatin, 5 insulin-like growth factor (somatomedins), luteinizing hormone releasing hormone (LHRH), tissue plasminogen activator (TPA), growth hormone releasing hormone (GHRH), oxytocin, estradiol, growth hormones, leuprolide acetate, factor VIII, interleukins such as interleukin-2, and analogs 10 thereof; analgesics such as fentanyl, sufentanil, butorphanol, buprenorphine, levorphanol, morphine, hydromorphone, hydcodone, oxymorphone, methadone, lidocaine, bupivacaine, diclofenac, naproxen, paverin, and analogs thereof; anti-migraine agents such as heparin, hirudin, and 15 analogs thereof; anti-coagulant agents such as scopolamine, ondansetron, domperidone, etoclopramide, and analogs thereof; cardiovascular agents, anti-hypertensive agents and vasodilators such as diltiazem, clonidine, nifedipine, verapamil, isosorbide-5-mononitrate, organic nitrates, agents 20 used in treatment of heart disorders and analogs thereof; sedatives such as benzodiazepines, phenothiazines and analogs thereof; narcotic antagonists such as naltrexone, naloxone and analogs thereof; chelating agents such as deferoxamine and analogs thereof; anti-diuretic agents such as 25 desmopressin, vasopressin and analogs thereof; anti-anginal agents such as nitroglycerine and analogs thereof; anti-neoplastics such as 5-fluorouracil, bleomycin and analogs thereof; prostaglandins and analogs thereof; and chemotherapy agents such as vincristine and analogs thereof.

30 Representative drugs also include but are not limited to antisense oligonucleotides, genes, gene correcting hybrid oligonucleotides, ribozymes, aptameric oligonucleotides, triple-helix forming oligonucleotides, inhibitors of signal transduction pathways, tyrosine kinase inhibitors and DNA 35 modifying agents. Drugs that can be used also include, without limitation, systems containing gene therapeutics, including viral systems for therapeutic gene delivery such as

adenovirus, adeno-associated virus, retroviruses, herpes simplex virus, sindbus virus, liposomes, cationic lipids, dendrimers, and enzymes. For instance, gene delivery viruses can be modified such that they express the targeting peptide 5 on the surface so as to permit targeted gene delivery.

In a preferred embodiment, a Therapeutic is therapeutically or prophylactically administered to a human patient.

Additional descriptions and sources of Therapeutics 10 that can be used according to the invention are found in various Sections herein.

#### 5.7. Therapeutic/Prophylactic Administration, Compositions and Formulations

15 The invention provides methods of treatment (and prophylaxis) by administration to a subject of an effective amount of a Therapeutic of the invention. In a preferred aspect, the Therapeutic is substantially purified. The subject is preferably an animal, including but not limited to 20 animals such as cows, pigs, horses, chickens, cats, dogs, etc., and is preferably a mammal, and most preferably a human.

As will be clear, any disease or disorder of 25 interest amenable to therapy or prophylaxis by providing a drug *in vivo* systemically or by targeting a drug *in vivo* to the GIT (by linkage to a GIT transport-receptor binding protein, derivative or analog of the invention) can be treated or prevented by administration of a Therapeutic of 30 the invention. Such diseases may include but are not limited to hypertension, diabetes, osteoporosis, hemophilia, anemia, cancer, migraine, and angina pectoris, to name but a few.

Any route of administration known in the art may be used, including but not limited to oral, nasal, topical, intravenous, intraperitoneal, intradermal, mucosal, 35 intrathecal, intramuscular, etc. Preferably, administration is oral; in such an embodiment the GIT-transport binding protein, derivative or analog of the invention acts

advantageously to facilitate transport of the therapeutic active agent through the lumen of the GIT into the systemic circulation.

The present invention also provides therapeutic compositions/formulations. In a specific embodiment of the invention, a GIT transport receptor-binding peptide or motif of interest is associated with a therapeutically or prophylactically active agent, preferably a drug or drug-containing nano- or microparticle. More preferably, the active agent is a drug encapsulating or drug loaded nano- or microparticle, such as a biodegradable nano- or microparticle, in which the peptide is physically adsorbed or coated or covalently bonded, such as directly linked or linked via a linking moiety, onto the surface of the nano- or microparticle. Alternatively, the peptide can form the nano- or microparticle itself or can be directly conjugated to the active agent. Such conjugations include fusion proteins in which a DNA sequence coding for the peptide is fused in-frame to the gene or cDNA coding for a therapeutic peptide or protein, such that the modified gene codes for a recombinant fusion protein in which the "targeting" peptide is fused to the therapeutic peptide or protein and where the "targeting" peptide increases the absorption of the fusion protein from the GIT. Preferably the particles range in size from 200-600 nm.

Thus, in a specific embodiment, a GIT transport-binding protein is bound to a slow-release (controlled release) device containing a drug. In a specific embodiment, polymeric materials can be used (see Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); Controlled Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, J. Macromol. Sci. Rev. Macromol. Chem. 23:61 (1983); see also Levy et al., 35 Science 228:190 (1985); During et al., Ann. Neurol. 25:351 (1989); Howard et al., J. Neurosurg. 71:105 (1989)).

The present invention also provides pharmaceutical compositions. Such compositions comprise a therapeutically effective amount of a Therapeutic, and a pharmaceutically acceptable carrier. In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, and more particularly in humans. The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the therapeutic is administered. Such pharmaceutical carriers can be sterile liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. Water is a preferred carrier when the pharmaceutical composition is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Suitable pharmaceutical excipients include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. These compositions can take the form of solutions, suspensions, emulsion, tablets, pills, capsules, powders, sustained-release formulations and the like. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulations can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, etc. Examples of suitable pharmaceutical carriers are described in "Remington's Pharmaceutical Sciences" by E.W. Martin. Such compositions will contain a therapeutically effective amount of the Therapeutic, preferably in purified

form, together with a suitable amount of carrier so as to provide the form for proper administration to the patient.

The Therapeutics of the invention can be formulated as neutral or salt forms. Pharmaceutically acceptable salts 5 include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, 10 triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The amount of the Therapeutic of the invention which will be effective in the treatment of a particular disorder or condition will depend on the nature of the 15 disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the 20 seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each patient's circumstances.

## 6. EXAMPLES

### 25 6.1. Selection of GIT Receptor Targets

The HPT1, hPEPT1, D2H, and hSI receptors were selected for cloning as GIT receptor targets based on several criteria, including: (1) expression on surface of epithelial cells in gastro-intestinal tract (GIT); (2) expression along 30 the length of small intestine (HPT1, hPEPT1, D2H); (3) expression locally at high concentration (hSI); (4) large putative extracellular domains facing into the lumen of the GIT; and (5) extracellular domains that permit easy access and bioadhesion by targeting particles.

35 The four recombinant receptor sites screened with the peptide libraries additionally have the following characteristics:

<u>Receptor</u>	<u>Characteristics</u>
D2H	Transport of neutral/basic amino acids; a transport activating protein for a range of amino acid translocases
5 hSI	Metabolism of sucrose and other sugars; represents 9% of brush border membrane protein in Jejunum
HPT1	di/tri peptide transporter or facilitator of peptide transport
hPEPT1	di/tri peptide transporter
10	Figures 1-4 (SEQ ID NOS:176, 178, 179, and 181, respectively) show the predicted amino acid sequences for hPEPT1, HPT1, hSI and D2H, respectively.

15 **6.2. Cloning of Extracellular Domain of Selected Receptor Site**

The following receptor domains were cloned and expressed as His-tag fusion proteins by standard techniques:

<u>Receptor</u>	<u>Domain (amino acid residues)</u>
20 hPEPT1 <sup>a</sup>	391-571
HPT1 <sup>b</sup>	29-273
hSI <sup>c</sup>	272-667
D2H <sup>d</sup>	387-685

25 <sup>a</sup> Liang et al., 1995, J. Biol. Chem. 270:6456-6463  
<sup>b</sup> Dantzig et al., 1994, Association of Intestinal Peptide Transport with a Protein Related to the Cadherin Superfamily  
<sup>c</sup> Chantret et al., Biochem. J. 285:915-923  
<sup>d</sup> Bertran et al., J. Biol. Chem. 268:14842-14949

30 The receptor proteins were expressed as His-tag fusion proteins and affinity purified under denaturing conditions, using urea or guanidine HCl, utilizing the pET His-tag metal chelate affinity for Ni-NTA Agarose (Hochuli, E., Purification of recombinant proteins with metal chelate adsorbent, Genetic Engineering, Principles and Methods (J.K. Setlow, ed.), Plenum Press, NY, Vol. 12 (1990), pp. 87-98).

### 6.3. Phage Libraries

Three phage DC8, D38, and DC43 libraries expressing N-terminal pIII fusions in M13 were used to identify peptides that bind to the GIT receptors. The D38 and DC43 libraries which are composed of 37 and 43 random amino acid domains, respectively, have been described previously (McConnell et al., 1995, Molecular Diversity, 1:165-176). The DC8 library is similar to the other two except that the random insert is 8 amino acids long flanked on each side by a cysteine residue (i.e., CX<sub>8</sub>C).

### 6.4. Biopanning

Three rounds of biopanning on the GIT receptors were performed generally by standard methods (McConnell et al., 1995, Molecular Diversity, 1:165-176), using a mixture of the DC8 ( $1 \times 10^{10}$  pfu), D38 and DC43 ( $1 \times 10^{11}$  pfu) phage libraries. After each round of panning the percentage of phage recovered was determined. Following the first two rounds of panning, the eluted phage were amplified overnight. Phage from the third pan were plated out and 100 plaques were picked, amplified overnight and screened in an ELISA assay for binding to the relevant receptor and BSA. After data analysis, phage clones were identified which had high absorbance in the ELISA assay and/or a good ratio of binding to target compared to binding to BSA. The Insulin Degrading Enzyme (IDE) and recombinant human tissue factor (hTF) were used as irrelevant controls. Several variations of the standard panning technique, discussed below, were used. Selection or panning methods followed one of two strategies. The first strategy involved panning the mixed libraries on the specific GIT receptor adsorbed to a solid surface. The second strategy panned the libraries twice against the GIT receptor and then against Caco-2 cells (Peterson and Mooseker, 1992, J. Cell Science 102:581-600), Selection methods are reflected in the clone nomenclature as described below:

S designates the clone was identified by binding to the hS1 receptor domain.

D designates the clone was identified by binding to the D2H receptor domain.

5 P designates the clone was identified by binding to the PEPT1 receptor domain.

H designates the clone was identified by binding to the HPT-1 receptor domain.

Phage designated Ni are from a solid phase band GIT 10 receptor pan that used the standard procedure with the addition of Ni-NTA Agarose (Qiagen, Chatsworth, CA).

Receptor coated plates were blocked with 0.5% BSA/PBS containing 160 $\mu$ l Ni-NTA agarose and libraries were panned in the presence of 50 $\mu$ l Ni-NTA agarose. The receptor proteins 15 were expressed as His-tag fusions. The His-tag has a high affinity for Ni-NTA Agarose. Blocking the plate and panning in the presence of Ni-NTA agarose minimized phage binding to the His-tag portion of the recombinant receptor.

Phage with the designation AX were eluted with acid 20 and Factor Xa. Phage were first eluted by standard acid elution then Factor Xa (New England Biolabs, Beverly, MA: 1 $\mu$ g protease in 300 $\mu$ l of 20mM Tris-HCL, 100mM NaCl, 2mM CaCl<sub>2</sub>) was added to the panning plate and incubated 2 hours. Phage from both elution methods were pooled together then plated.

25 Phage with the designation AB were eluted with acid and base. Phage were eluted first by standard acid elution then 100mM triethylamine pH 12.1 was added to the panning plate for 10 minutes. Phage from both elution methods were pooled together then plated.

30 C designates panning on receptor followed by Caco-2 cells. First and second round pans were performed on the receptor and the third round pan was on snapwells of Caco-2 cells. DCX11, DCX8 and DCX33 were identified by two pans on D2H receptor, third pan on Caco-2 cells. The third round 35 Factor Xa eluate from the Caco-2 cells was screened by ELISA on D2H, BSA and fixed Caco-2 cells. For HCA3 the first two rounds of panning were performed on the HPT-1 receptor and

the third pan was on monolayers cultured on snapwells of Caco-2 cells.

Phage designated 5PAX were carried through five rounds of panning after which a number of phage were 5 sequenced prior to screening by ELISA.

#### 6.5. Sequencing of Selected Phage

The amino acid sequence of phage inserts demonstrating a good ratio of binding to receptor domains 10 and/or Caco-2 cells over background BSA binding were deduced from the nucleotide sequence obtained by sequencing (Sequenase®, U.S. Biochemical Corp., Cleveland, OH) both DNA strands of the appropriate region in the viral genome. The third round acid eluate was screened by ELISA on HPT-1, BSA 15 and Caco-2 fixed cells. Phage designated 5PAX were carried through five rounds of panning after which a number of phages were sequenced prior to screening by ELISA.

One well of a 24 well plate was coated with 10  $\mu$ g/ml of GIT receptor and the plate was incubated overnight 20 at 4°C. The plate was blocked with 0.5 BSA-PBS for one hour. A mixture of the DC8, D38 and DC43 phage libraries was added to the plate and the plate was incubated for 2 to 3 hours at room temperature on a rotator. After washing the well 10 times with 1% BSA plus 0.05% Tween 20 in PBS, the well was 25 eluted with 0.05M glycine, pH2. The phage was then eluted with 0.2M NaPO<sub>4</sub>. The eluted phage was titered on agar plates; the remaining phage was amplified overnight. The next day the amplified phage was added to a second coated plate and the panning procedure was repeated as described above. The 30 35 eluted phage from the second pan as well as the amplified phage from the first pan was titered on agar plates. Following amplification overnight of the phage from the second pan, the panning procedure was repeated as described above. The phage eluted from the third pan and the amplified phage from the second pan were then titered overnight on agar plates. Isolated phage colonies were amplified overnight prior to use in an ELISA assay.

#### 6.6. Receptor ELISA Procedure

96 well plates were coated overnight with GIT receptor, BSA and, optionally, IDE (insulin degrading enzyme, an irrelevant His-fusion protein) or hTF. The plates were 5 blocked for one hour with 0.5% BSA-PBS. After clarification, the amplified phage were diluted 1:100 in 1% BSA plus 0.05% Tween 20 in PBS and added to the plates. Following incubation of the plates on a rotator for 1 to 2 hours, the plates were washed 5 times with 1% BSA plus 0.05% Tween 20 in 10 PBS. Dilute anti-M13-HRP conjugate (anti-M13 antibody linked to horse radish peroxidase (HRP)) was added to all the wells and the plate was incubated for one hour on a rotator. After the plates were washed 5 times, as described above, TMB substrate was added to the wells. The plates were read at 15 650nm absorbance.

#### RECEPTOR ELISA RESULTS:

Below are the results of ELISA assays which assessed the binding of phage panned on the hSI receptor to 20 microtiter plates coated with hSI and BSA. Table 1 shows the OD results as well as the ratio of hSI to BSA binding.

25

30

35

Table 1

PHAGE	hSI	BSA	hSI/BSA
S15	0.478	0.053	9
S21	0.845	0.092	9
S22	0.399	0.061	7
SNI10	0.57	0.051	11
SNI28	0.942	0.113	8
SNI34	0.761	0.115	7
SNI38	0.466	0.076	6
SNI45	0.518	0.056	9
SNIAX2	0.383	0.065	6
SNIAX6	0.369	0.056	7
SNIAX8	0.342	0.068	5
BLANK	0.063	0.042	2

Below are the results of an ELISA which assessed the binding of phage panned on the D2H receptor to microtiter plates coated with D2H and BSA. Table 2 shows the OD results as well as the ratio of D2H to BSA binding.

Table 2

Phage	D2H	BSA	D2H/BSA
DAB3	0.406	0.072	6
DAB7	0.702	0.09	8
DAB10	0.644	0.153	4
DAB18	0.467	0.085	5
DAB24	1.801	0.441	4
DAB30	0.704	0.121	6
DAX15	0.391	0.101	4
DAX23	0.698	0.153	5
DAX24	0.591	0.118	5
DAX27	1.577	0.424	4
BLANK	0.038	0.037	1

Below are the results of an ELISA which assessed the binding of phage panned for two rounds on the D2H receptor followed by a third round pan on Caco-2 snapwells. Binding to fixed Caco-2 cells, D2H and BSA was examined.

Table 3 shows the OD results as well as the ratio of D2H to BSA binding.

Table 3

PHAGE	Caco-2	D2H	BSA	D2H/BSA
DCX8	0.498	0.163	0.063	3
DCX11	0.224	0.222	0.071	3
DCX26	0.114	0.956	0.213	4
DCX33	0.164	0.616	0.103	6
DCX36	0.149	0.293	0.064	5
DCX39	0.121	0.299	0.066	5
DCX42	0.308	0.158	0.065	2
DCX45	0.147	0.336	0.075	4
Blank	0.065	0.043	0.04	1

Below are the results of an ELISA which assessed the binding of phage panned on the hPEPT1 receptor to hPEPT1 and BSA. Table 4 shows the OD results as well as the ratio of hPEPT1 to BSA binding.

Table 4

PHAGE	hPEPT1	BSA	PEPT1/BSA
PAX9	0.312	0.079	4
PAX14	1.102	0.139	8
PAX15	0.301	0.079	4
PAX16	0.648	0.171	4
PAX17	0.514	0.095	5
PAX18	0.416	0.087	5
PAX35	0.474	0.065	7
PAX38	0.292	0.064	5
PAX40	0.461	0.076	6
PAX43	0.345	0.069	5
PAX45	0.419	0.081	5
PAX46	0.429	0.077	6
P31	0.807	0.075	11
P90	1.117	0.107	9
5PAX3	0.173	0.04	4
5PAX5	0.15	0.036	4
5PAX7	0.171	0.037	5
5PAX12	0.227	0.04	6
Blank	0.102	0.039	3

Table 5 shows the results of an ELISA which assessed the binding of phage panned on the HPT-1 receptor to HPT-1 and BSA. The table shows the OD results as well as the ratio of HPT-1 to BSA binding.

5

Table 5

PHAGE	HPT1	BSA	HPT/BSA
HAX9	0.382	0.075	5
HAX40	0.991	0.065	15
HAX42	0.32	0.071	5

10

Table 6 shows the results of an ELISA which assessed the binding of phage panned for two rounds on the HPT-1 receptor followed by a third round pan on Caco-2 snapwells. Binding to fixed Caco-2 cells, HPT-1 and BSA was examined. The table shows the OD results as well as the ratio of HPT-1 to BSA binding.

15

Table 6

20

PHAGE	Caco-2	HPT1	BSA	HPT1/BSA
HCA3	0.406	0.048	0.038	1

#### CELL ELISA PROCEDURE

Phage ELISA was used as described above with the following changes. Diluent and wash buffer was PBS containing 1%BSA and 0.05% Tween 20 and plates were washed five times at each wash step. Supernatant of infected bacterial cultures was diluted 1:100 and incubated with protein coated plates for 2-3 hours with mild agitation.

Anti-M13 Horseradish peroxidase (HRP) conjugate (Pharmacia, Piscataway, NJ) was diluted 1:8000.

Fixed Caco-2, C2BBe1, and A431 cell plates were prepared by growing cells on tissue culture treated microtiter plates. When cells were confluent, plates were fixed with 10% formaldehyde, washed twice with PBS and stored with 0.5%BSA-PBS at -20°C. On the day of the assay, thawed

plates were treated with PBS containing 0.1% phenylhydrazine for one hour at 37°C followed by two PBS washes and blocking for one hour with 0.5% BSA-PBS. The standard ELISA procedure was followed at this point.

5 Phage which showed specificity to a GIT receptor was further characterized by ELISA on a variety of recombinant proteins. Phage which continued to exhibit GIT receptor specificity was sequenced.

10

Table 7

TARGET BINDING PHAGE INSERT SEQUENCES:

	SEQ.
hSI	ID. NO.
S15	1 RSGAYESPDRGGRSYVGCGNIGRKHNWLGLRTASPACWD
S21	2 SPRSFVPPVSRHESFGISNYLGCGYRTCISGTMTKSSPIYPRHS
15 S22	3 SSSSDWGGVPGKVVRRERFKGRGCGISITSVLTGKPNPCPEPKAA
SNI10	4 RVGQCTDSDVRRPWARSCAHQCGAGTRNSHGCITRPLRQASAH
SNI28	5 SHSGGMNRAYGDVFRRELDRWNATSHHTRPTPQLPRGPN
SNI34	6 SPCGGSWGRFMQGGLFGGRTDGCAGHRNRTSASLEPPSSDY
SNI38	7 RGAADQRRGWSENGLPRVGWDAIAHNSYTFSTSRRPRPP
20 SNI45	8 SGGEVSSWGRVNNDLCARVSWTGCCTARSARTDNKGFLPKHSSLR
SNIAX2	9 SDSDGDHYGLRGGVRCSLRDRGCGLALSTVHAGPPSFYPKLSSP
SNIAX4	10 RSLGNYGVTGTVDTVLPMPGHANHLGVSSASSSDPPRR
SNIAX6	11 RTTAKGCLLGSFGVLSGCSFTPTSPPPHLGYPPHSVN
SNIAX8	12 SPKLSSVGVMTKVTELPTEGPNAISIPISATLGPRNPLR
25	
	<u>D2H</u>
DAB3	13 RWCGAELCNSVTKKFRPGWRDHANPSTHHRTPPPSQSSP
DAB7	14 RWCGADDPCGASRWRGGNSLFGCGLRCSSAQSTPSGRIHSTSTS
DAB10	15 SKSGEGGDSSRGETGWARVRSHAMTAGRFRWYNQLPSDR
30 DAB18	16 RSSANNCEWKSDWMRRACIARYANSSGPARAVDTKAAP
DAB24	17 SKWSWSSRWGSPQDKVEKTRAGCGGSPSSTNCHPYTFAPPPQAG
DAB30	18 SGFWEFSRGLWDGENRKSVRSGCGFRGSSAQGPCPVTPATIDKH
DAX15	19 SESGRCRSVSRWMTTWQTQKGGCGSNVSRGSPLDPSHQTGHATT
DAX23	20 REWRFAGPPLDLWAGPSLPSFNASSHPRALRTYWSQRPR
35 DAX24	21 RMEDIKNSGWRDSCRWGDLRPGCGSRQWYPSNMRSSRDYPAGGH
DAX27	22 SHPWYRHWNHGDFSGSGQSRHTPPESPHGRPNATI

DCX8	23	RYKHDIGCDAGVDKKSSSVRGCGAHSSPPRAGRGPRTMVSRL
DCX11	24	SQGSKQCMQYRTGRLTVGSEYCGMNPARHATPAYPARLLPRYR
DCX26	25	SGRTTSEISGLWGWDRLSGYWGNTLRPNYIPYRQATNRHRYT
DCX33	26	RWNWTVLPATGGHYWTRSTDYHAINNHRPSIPHQHPTPI
5 DCX36	27	SWSSWNWSSKTTRLGDRATREGCGPSQSDGCPYNGRLTTVKPRT
DCX39	28	SGSLNAWQPRSWVGGAFRSHANNLNPKPTMVTRHPT
DCX42	29	RYSGLSPRDNGPACSQEATLEGCGAQRLMSTRRKGRNSRPGWTL
DCX45	30	SVGNDKTSRPVSYGRVSDLWNASLMPKRTPSSKRHDDG
10 <u>hPEPT1</u>		
PAX9	31	RWPSVGYKGNGSDTIDVHSNDASTKRSЛИYNHRRPLFP
PAX14	32	RTFENDGLGVGRSIQKKSDRwyASHNIRSHFASMSPAGK
PAX15	33	SYCRVKGGEGGGHTDSNLARSGCGKVARTSRLQHINPRATPPSR
PAX16	34	SWTRWGKHTHGGFVNKSPPGKNATSPYTDALQPLSDQGPP
15 PAX17	35	SQVDSFRNSFRWYEPSRALCHCGKRDSTTRIHNSPSDSYPTR
PAX18	36	SFLRFQSPRFEDYSRTISRLRNATNPSNVSDAHNNRALA
PAX35	37	RSITDGGINEVDLSSVSNVLENANSHRAYRKHRPTLKRP
PAX38	38	SSKVSSPRDPTVPRKGGNVYGCGRSSARMPTSALSSITKCYT
PAX40	39	RASTQGGRGVAPEFGASVLGRGCGSATYYTNSTSCDKAMGHNYS
20 PAX43	40	RWCEKHKFTAARCSAGAGFERDASRPPQPAHRDNTNRNA
PAX45	41	SFQVYPDHGLERHALDGTGPLYAMPGRWIRARPQNRDRQ
PAX46	42	SRCTDNEQCPDTGTRSRSVSNARYFSSRLLKTHAPHRP
P31	43	SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRRHP
P90	44	SSADAEKCAGSLLWWGRQNNNSCGSPTKKHLKHRNRSQTSSSH
25 5PAX3	45	RPKNVADAYSSQDGAAEETSHASNAARKSPKHKPLRRP
5PAX5	46	RGSTGTAGGERSGVNLHTRDNASGSGFKPWPSNRGHK
5PAX7	47	RWGWERSPSDYDSDMDLGARRYATRTHRAPPRVLKALP
5PAX12	48	RGWKCEGSQAAYGDKDIGRSRGCGSITKNNTNHAHPHSGAVAKI
30 <u>HPT-1</u>		
HAX9	49	SREEANWDGYKREMSHRSRWDATHLSRPRPANSQDPM
HAX35	50	EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPRK
HAX40	51	REFAERRLWGDDLSWRDAEGCGPTPSNRAVKHRKPRPRSPAL
HAX42	52	SDHALGTNLRSNAKEPGDYNCCGNGNSTGRKVFNRRRPSAIPT
35 HCA3	53	RHISEYSFANSHLMGGESKRKGCGINGSFSPTCPRSPTPAFRRT
H40	54	SRESGMWGSWRGHRLNSTGGNAMNASLPPDPPVSTP
PAX2	55	STPPSREAYSRPYSVSDSDTNAKHSSHNRRLTRSRPN

Table 8

DNA Sequences for Clones used in *in vivo* Pan

## S15 (SEQ ID NO: 56)

5 TCTCACTCCTCGAGATCCGGCGCTTATGAGAGTCCGGATGGTCGGGGGGTCGGAGCTATG  
 TGGGGGGCGGGGGTGGNTGGTAACATTGGTCGGAAGCATAACCTGTGGGGCTGCGTAC  
 CGCGTCGCCGGCCTGCTGGACTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## S21 (SEQ ID NO: 57)

10 TCTCACTCCTCGAGTCCTCGCTCTTCTGGCCCGTTGTGTCGGCATGAGTCGTTGGGA  
 TCTCTAACTATTGGGNTGGTTATCGTACATGTATCTCCGGCACGATGACTAAGTCTAG  
 CCCGATTACCCCTCGGCATTGTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## S22 (SEQ ID NO: 58)

TCTCACTCCTCGAGTAGTAGCTCGATTGGGTGGTGTGCCCTGGGAAGGTGGTAGGGAGC  
 GCTTAAGGGGCGCGGGTGTGGTATTCCATCACCTCCGTGCTACTGGGAAGCCAAATCC  
 GTGTCCGGAGCCTAAGCGGCCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

15 SNI 10 (SEQ ID NO: 59)

TCTCACTCCTCGAGAGTTGCCAGTGCACGGATTCTGATGTGCGGCCGTCTGGGCCAGGT  
 CTTGCGCTCATCAGGGTTGTGGTGCAGGCACTCGCAACTCGCACGGCTGCATACCCGTCC  
 TCTCCGCCAGGCTAGCGCTATTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

20 SNI 28 (SEQ ID NO: 60)

TCTCACTCCTCGAGCCACTCCGGTGGTATGAATAGGGCTACGGGGATGTGTTAGGGAGC  
 TTCGTGATCGGTGGAACGCCACTTCCCACCACACTGCCACCCCTCAGCTCCCCCGTGG  
 GCCTAATTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## SNI 34 (SEQ ID NO: 61)

25 TCTCACTCCTCGAGTCCGTGCGGGGGTCGTGGGGCGTTTATGCAGGGTGGCCTTTCG  
 GCGGTAGGACTGATGGTTGTGGTGCCTAGAAACCGCAGTCTCGTGTAGAGCCCC  
 GAGCAGCGACTACTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## SNI 38 (SEQ ID NO: 62)

30 TCTCACTCCTCGAGGGGCGCCGCGATCAGCGCGGGGTGGTCCGAGAACCTGGGGTTGC  
 CTAGGGTGGGGTGGGACGCCATCGCTACAATAGCTACGTTACCTCGCGCCGCCGCG  
 CCCCCCTCTAGA

## SNI 45 (SEQ ID NO: 63)

TCTCACTCCTCGAGCGGTGGGAGGTCAGCTCTGGGCGCGTGAATGACCTCTCGCGCTA  
 GGGTGAGTTGGACTGGTTGTGGTACTGCTCGTCCGCGCGTACCGACAACAAAGGTTCT  
 TCCTAACGCACTCGTACTCCGCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

35 SNI AX2 (SEQ ID NO: 64)

TCTCACTCCTCGAGTGATAGTGACGGGGATCATTATGGGCTCGGGGGGGGTGCGTTGTT  
 CGCTTCGTGATAGGGTTGTGGTCTGGCCCTGTCACCGTCCATGCTGGTCCCCCTCTT  
 TTACCCCAAGCTCTCCAGCCCCCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**SNI AX4 (SEQ ID NO: 65)**

5 TCTCACTCCTCGAGGAGCTGGTAATTATGGCGTCACCGGGACTGTGGACGTGACGGTT  
 TGCCCATGCCTGGCCACGCCAACACCTGGTGTCTCCCTCGCCTAGCTCTGATCCTCC  
 GCGCGCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**SNI AX6 (SEQ ID NO: 66)**

10 TCTCACTCCTCGAGAACTACGACGGCTAAGGGGTGCTTCTCGGAAGCTCGGCGTTCTTA  
 GTGGGTGCTCATTACGCCAACCTCTCCACCGCCCCACCTAGGATAACCCCCCCCACTCCTG  
 CAATTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**SNI AX8 (SEQ ID NO: 67)**

15 TCTCACTCCTCGAGCCGAAGTTGTCCAGCGTGGGTGTTATGACTAAGGTACGGAGCTGC  
 CCACGGAGGGGCCTAACGCCATTAGTATTCCGATCTCCCGACCCCTCGGCCCCGCAACCC  
 GCTCCGCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

15 **DAB3 (SEQ ID NO: 68)**

TCTCACTCCTCGAGGGTGGTGCAGCGCTGAGCTGTGCAACTCGGTGACTAAGAAGTTCGCC  
 CGGGCTGGCGGGATCACGCCATCCCTCCACCCATCATCGTACTCCCCGCCCAGCCAGTC  
 CAGCCCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

20 **DAB7 (SEQ ID NO: 69)**

TCTCACTCCTCGAGGGTGGTGCAGCGCTGATGACCCGTGTGGTGCCAGTCGTTGGCGGGGG  
 GCAACAGCTTGTGGTGTGGTCTTCGTTGTAGTGCAGGAGCACCCGAGTGGCAG  
 GATCCATTCCACTTCGACCAGCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**DAB10 (SEQ ID NO: 70)**

25 TCTCACTCCTCGAGTAAGTCGGGGAGGGGGGTGACAGTAGCAGGGCGAGACGGGCTGG  
 CGAGGGTTCGGTCTCACGCCATGACTGCTGGCCCTTCGGTGGTACAACCAGTTGCCCTC  
 TGATCGGTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**DAB18 (SEQ ID NO: 71)**

30 TCTCACTCCTCGAGGTGAGCGCCAATAATTGCGAGTGGAAAGTCTGATTGGATGCGCAGGG  
 CCTGTATTGCTCGTTACGCCAACAGTTGGCCCGCCCGCCGCGACACTAAGGCCGC  
 GCCCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**DAB24 (SEQ ID NO: 72)**

TCTCACTCCTCGAGTAAGTGGTCGTGGAGTTCGAGGTGGGGCTCCCCGCAGGATAAGGTTG  
 AGAAGACCAGGGCGGGTTGTGGTGGTAGTCCCAGCAGCACCAATTGTACCCCTACACCTT  
 TGCCCCCCCCCGCAAGCCGGCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

35

## DAB30 (SEQ ID NO: 73)

TCTCACTCCTCGAGTGGGTTCTGGGAGTTAGCAGGGGGCTTGGGATGGGAGAACGTA  
 AGAGTGTCCGGTCGGGTTGTGGTTCTGGCTCTGCTCAGGGCCGTGTCGGTCAC  
 GCCTGCCACCATTGACAAACACTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

5

## DAX15 (SEQ ID NO: 74)

TCTCACTCCTCGAGTGAGAGCGGGCGGTGCCGTAGCGTAGGCCGGTGGATGACGACGTGGC  
 AGACGCAGAAGGGCGGTTGTGGTTCAATGTTCCCGCGGTTGCCCTCGACCCCTCTCA  
 CCAGACCGGGCATGCCACTACTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

10

## DAX23 (SEQ ID NO: 75)

TCTCACTCCTCGAGGGAGTGGAGGTTGCCGGGCCGCTGGACCTGTGGGCGGGTCCGA  
 GCTTGCCTCTTTAACGCCAGTCCACCCCTCGGCCCTGCGCACCTATTGGTCCCAGCG  
 GCCCGCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## DAX24 (SEQ ID NO: 76)

15

TCTCACTCCTCGAGGATGGAGGACATCAAGAACACTGGGGTGGAGGGACTCTTGTAGGTGGG  
 GTGACCTGAGGCCTGGTTGTGGTAGCCGCAGTGGTACCCCTCGAATATGGTTCTAGCAG  
 AGATTACCCCGGGGGCCACTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## DAX27 (SEQ ID NO: 77)

20

TCTCACTCCTCGAGTCATCCGTGGTACAGGCATTGGAACCATGGTACTTCTCTGGTTCGG  
 GCCAGTCACGCCACACCCCGCCGGAGAGCCCCCACCCGGCCGCCCTAATGCCACCATTTC  
 TAGAATCGAAGGTCGCGCTAGACCTTCGAG

## DCX8 (SEQ ID NO: 78)

25

TCTCACTCCTCGAGATATAAGCACGATATCGGTGCGATGCTGGGTTGACAAGAAGTCGT  
 CGTCTGTGCGTGGTGTGGTCTAGTNGTCGCCACCCCGCGCCGGCGTGGTCTCG  
 CGGCACGATGGTTAGCAGGCTTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## DCX11 (SEQ ID NO: 79)

30

TCTCACTCCTCGAGTCAGGGCTCCAAGCAGTGTATGCAGTACCGCACCGGTGTTGACGG  
 TGGGGTCTGAGTATGGTTGTGGTATGAACCCGCCCATGCCACGCCGCTTATCCGGC  
 GCGCCTGCTGCCACGCTATCGCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## DCX26 (SEQ ID NO: 80)

35

TCTCACTCCTCGAGTGGCGGACTACTAGTGAGATTCTGGCTCTGGGGTTGGGTGACG  
 ACCGGAGCGGTTATGGTTGGGTAACACGCTCCGCCCAACTACATCCCTATAGGCAGGC  
 GACGAACAGGCATCGTTATACGTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## DCX33 (SEQ ID NO: 81)

40

TCTCACTCCTCGAGGTGGAATTGGACTGTCTTGGCCGCCACTGGGGCCATTACTGGACGC  
 GTTCGACGGACTATCACGCCATTAACAAATCACAGGGCGAGCATCCCCACCAGCATCCGAC  
 CCCTATCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## DCX36 (SEQ ID NO: 82)

TCTCACTCCTCGAGTTGGTCGTGGATTGGAGCTCTAAGACTACTCGTCTGGCGACA  
 GGGCGACTCGGGAGGGTTGTGGTCCCAGCCAGTCTGATGGCTGTCTTATAACGGCCGCT  
 TACGACCGTCAAGCCTCGCACGTCTAGAATCGAAGGTCGCGTAGACCTTCGAGA

5

## DCX39 (SEQ ID NO: 83)

TCTCACTCCTCGAGTGGTAGTTGAACGCATGGCAACCGCGGTATGGTGGGGCGCGT  
 TCCGGTCACACGCCAACAAATAACTGAAACCCCAAGCCCACCATGGTTACTNGTCACCCTAC  
 CTCTAGAATCGAAGGTCGCGTAGACCTTCGAGA

10

## DCX42 (SEQ ID NO: 84)

TCTCACTCCTCGAGGTATTGGGTTTGTCCCCGGGGACAACGGTCCCCTGTAGTCAGG  
 AGGCTACCTTGGAGGGTTGTGGTGCAGAGGCTGATGTCACCCGTGCAAGGGCCGCAA  
 CTCCCGCCCCGGGTGGACGCTCTAGAATCGAAGGTCGCGTAGACCTTCGAGA

## DCX45 (SEQ ID NO: 85)

15

TCTCACTCCTCGAGCGTGGGAATGATAAGACTAGCAGGCCGGTTCCCTACGGCGCG  
 TTAGTGATCTGTGGAACGCCAGCTTGATGCCAACCGTACTCCAGCTCGAACGCCACGA  
 TGATGGCTCTAGAATCGAAGGTCGCGTAGACCTTCGAGA

## PAX2 (SEQ ID NO: 86)

20

TCTCACTCCTCGAGTACTCCCCCAGTAGGGAGGCAGTATAGTAGGCCCTATAGTGTGATA  
 GCGATTGGATACGAACGCCAACAGCACAGCTCCACAACGCCGNTGGACGCCAGCG  
 CCCGAACCTAGAATCGAAGGTCGCGTAGACCTTCGAGA

## PAX9 (SEQ ID NO: 87)

25

TCTCACTCCTCGAGATGGCTAGTGTGGTTACAAGGTAATGGCAGTGACACTATTGATG  
 TTCACAGCAATGACGCCAGTACTAAGAGGTCCCTCATCTATAACCACGCCGCCCCNTCTT  
 TCCCTCTAGAATCGAAGGTCGCGTAGACCTTCGAGA

## PAX14 (SEQ ID NO: 88)

30

TCTCACTCCTCGAGAACGTTGAGAACGACGGGCTGGCGTGGCCGGTCTATTCAAGAAGA  
 AGTCGGATAGGTGGTACGCCAGCCACACATTGCTAGCCATTGCGTCCATGTCTCCGC  
 TGGTAAGTCTAGAATCGAAGGTCGCGTAGACCTTCGAGA

## PAX15 (SEQ ID NO: 89)

35

TCTCACTCCTCGAGCTATTGTCGGGTTAAGGGTGGTGGGGAGGGGGGGCATACGGATTCCA  
 ATCTGGCTAGGTGGGTTGTGGTAAGGTGGCCAGGACCAGCAGGCTTCAGCATATCAACCC  
 GCGCGCTACCCCCCCCCTCCCGTCTAGAATCGAAGGTC

## PAX16 (SEQ ID NO: 90)

40

TCTCACTCCTCGAGTTGGACTCGGTGGGCAAGCACANTCATGGGGGTTGTGAACAAGT  
 CTCCCCCTGGGAAGAACGCCACGAGCCCTACACCGACGCCAGCTGCCAGTGATCAGGG  
 TCCTCCCTCTAGAATCGAAGGTCGCGTAGACCTTCGAGA

## PAX17 (SEQ ID NO: 91)

TCTCACTCCTCGAGTCAGGTTGATTGTTCTGTAATAGCTTCGGTGGTATGAGCCGAGCA  
GGGCTCTGTGCCATGGTTGTTGAAGCGCGACACCTCCACCACTCGTATCCACAATAGCCC  
CAGCGACTCCTATCCTACACGCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

5

## PAX18 (SEQ ID NO: 92)

TCTCACTCCTCGAGCTTTGCGGTTCCAGAGTCCGAGGTTGAGGATTACAGTAGGACGA  
TCTNTCGGTTGCGAACGCCACGAACCCGAGTAATGTCTCCGATGCGCACAATAACCGGGC  
CTTGGCCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## 10 PAX35 (SEQ ID NO: 93)

TCTCACTCCTCGAGGAGCATACCGACGGGGCATCAATGAGGTGGACCTGAGTAGTGTGT  
CGAACGTTCTGAGAACGCCACTCGCATAGGGCCTACAGGAAGCATGCCGACCTTGAA  
GCGTCCTTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## PAX38 (SEQ ID NO: 94)

15 TCTCACTCCTCGAGTTCGAAGGTGAGCGAGCCCGAGGGATCCGACGGTCCCAGGGAAAGGGCG  
GCAATGTTGATTATGGTTGTTGTCACAGGTCTCCGCCGGATGCTTACCTCCGCTCTGTC  
GTCGATCACGAAGTGCTACACTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## PAX40 (SEQ ID NO: 95)

20 TCTCACTCCTCGAGAGGCCAGTANGCAGGGCGCCGGGTGTTGCCCTGAGTTGGGGCGA  
GCGTTTGGGTNGTGGTTGTTGTCACAGGTCTCCGCCGGATGCTTACCTCCACCAAGCTGCAAGGA  
TGCTATGGGCCACAACTACTCGTCTAGAATCGAAGGTCGCGNTAGACCTTCGAGA

## PAX43 (SEQ ID NO: 96)

25 TCTCACTCCTCGAGATGGTGCAGAGAAGCACAAGTTACGGCTGCGGTTGCAGCGGGGG  
CGGGTTTGAGAGGGANGCCAGCCGTCCGCCAGCCTGCCACCGGATAATACCAACCG  
TAATGCNTNTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

25

## PAX45 (SEQ ID NO: 97)

TCTCACTCCTCGAGTTTCAGGTGTACCCGGACCATGGTCTGGAGAGGCATGCTTGACG  
GGACGGGTCCGCTTACGCCATGCCGGCGCTGGATTAGGGCGCGTCCGAGAACAGGGA  
CCGCCAGTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

30

## PAX46 (SEQ ID NO: 98)

TCTCACTCCTCGAGCAGGTGTACGGACAACGAGCAGTGCCTGAGGGATACCGGGANTAGGTCTC  
GTTCCGTTAGTAACGCCAGGTACTTTGAGCAGGTTGCTCAAGACTCACGCCCTCATCG  
CCCTTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## P31 (SEQ ID NO: 99)

35 TCTCACTCCTCGAGTGCCAGGGATAGCGGGCTGCGGAGGGATGGTCCCGCGCGTCCGGT  
TGAACGGGGTTGAGAACGCCAACACTAGGAAGTCCCTCCCGAGTAACCCGCGGGTAGGCG  
CCATCCCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**P90 (SEQ ID NO: 100)**

TCTCACTCCTCGAGTTCCGCCGATGCGGAGAAGTGTGCGGGCAGTCTGTTGGTGGGTA  
 GGCAGAACAACTCCGGTTGTGGTCCGCCCACGAAGAACATCTGAAGCACCGCAATCGCAG  
 TCAGACCTCCTCTCGTCCCACTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**5 PAX3 (SEQ ID NO: 101)**

TCTCACTCCTCGAGAGACCGAAGAACGTGGCCGATGCTTATTGCTCTCAGGACGGGGCGGGCGG  
 CCGAGGAGACGTCTCACGCCAGTAATGCCGCCGGAAAGTCCCCTAACGACAAGCCCTTGAG  
 CGGGCTTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**5PAX5 (SEQ ID NO: 102)**

**10** TCTCACTCCTCGAGAGAGGAGTACGGGGACGGCCGGCGAGCGTTCCGGGTGCTCAACC  
 TGCACACCAGGGATAACGCCAGCGGAGCGGTTCAAACCGTGGTACCCCTCGAATCGGGG  
 TCACAAGTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**5PAX7 (SEQ ID NO: 103)**

**15** TCTCACTCCTCGAGGTGGGGTGGGAGAGGGAGTCCGTCCGACTACGATTCTGATATGGACT  
 TGGGGGCGAGGAGGTACGCCACCCGCACCCACCGCGGCCCTCGCGTCTGAAGGCTCC  
 CCTGCCCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**5PAX12 (SEQ ID NO: 104)**

**20** TCTCACTCCTCGAGGCACTGGAAGTGCAGGGCTCTCAGGCTGCCTACGGGACAAGGATA  
 TCGGGAGGTCCAGGGTTGTGGTCCATTACAAAGAATAACACTAATCACGCCATCCTAG  
 CCACGGCGCCGTTGCTAAGATCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**HAX9 (SEQ ID NO: 105)**

TCTCACTCCTCGAGCCCGAGGAGGGCGAACTGGGACGGCTATAAGAGGGAGATGAGCCACC  
 GGAGTCGTTTGGGACGCCACCCACCTGCCCCCTGCCGCCGCTAACTCTGGTGA  
 CCCTAACTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**25 HAX40 (SEQ ID NO: 106)**

TCTCACTCNTCGAGAGAGTTCGCGGAGAGGGAGGTTGTGGGGGTGTGATGACCTGAGTTGGC  
 GTCTCGACGCCAGGGTTGTGGTCCCCTCCGAGCAATCGGGCCGTCAAGCATCGCAAGGCC  
 CCGCCCACGCTCCCCCGACTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**HAX42 (SEQ ID NO: 107)**

**30** TCTCACTCNTNGAGTGTACCGCGTTGGGACGAATCTGAGGTCTGACAATGCCAAGGAGC  
 CGGGTGATTACAACGTGTGGTAACGGGAACCTTACCGGGCGAAAGGTTTTAACCGTAG  
 CGGCCCTCCGCCATCCCCANTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

**HCA3 (SEQ ID NO: 108)**

**35** TCTCACTCCTCGAGGCATATTCTGAGTATAGCTTGCAGATTCCCACTTGATGGGTGGCG  
 AGTCCAAGCGGAAGGGTTGTGGTATTAACGGCTCTTCTCCCACCTGTCCCCGCTCCCC  
 CACCCCGCCCTCCGCCACCTCTAGAATCGAAGGTCGCGCTAGACCTTCGAGA

## H40 (SEQ ID NO: 109)

TCTCACTCCTCGAGCCGGAGAGCGGGATGTGGGTAGTTGGTGGCGTGGTCACAGGTTGA  
 ATTCCACGGGGGTAAACGCCAACATGAATGCTAGTCTGCCCGACCCCCCTGTTCCAC  
 TCCGTCTAGAATCGAAGGTGCGCTAGACCTTCGAG

5 Peptide Motifs

By comparison of the amino acid sequences of the clones binding GIT receptors, certain sequence similarities or "motifs" were recognized. These motifs can often represent the part of the sequence that is important for 10 binding to the target. Table 9 identifies regions of sequence similarity or sequence motifs (in boldface) that were identified among GIT binding peptides (corresponding SEQ ID NOS. are shown in Table 7).

15

Table 9

PEPT-1	
HPT1	
P31	SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRRHP
PAX9	RWPSPVGKGNNSDTIDVHSNDASTKRSLIYNHRRPLFP
HAX42	SDHALGTNLRSDNAKEPGDYNCCGNGNSTGRK- <b>VFNRRRPSAIPT</b>
PAX2	STPPSREAYSRPYSVDSDDTNAKHSSHNRRLTRSRPN
20	
hSI	
SNI10	RVGQCTDSDVRRPWARSCAHQCGAGTRNSHGCITRPLRQASAH
SNI38	RGAADQRRGWSENGLPVGWDAIAHNSYTFTSRRPRPP
S15	RSGAYESPDRGGRSYVGGGGCGNIGRKHNLWGLRTASPACWD
SNI34	SPCGGSWGRFMQGGLFGGRTDGCGAHRNRTSASLEPPSSDY
25	
D2H	
DAB10	SKSGEGGDSSRGETGWARVRSHAMTAGRFRWYNQLPSDR
DAB30	SGFWEFSRGLWDGENRKSVRSGCGFRGSSAQGPCPVTPATIDKH
DCX8	RYKHDIGCDAGVDKSSSVRGCG- <b>AHSSPPRAGRGPRTMVSRL</b>

Phage Binding to Caco-2 Cells

30

Phage expressing presumed GIT binding peptide inserts were also assayed by ELISA on fixed Caco-2 or C2BBe1 cells as follows. Cells were plated at  $1 \times 10^5$  cells/well on 100  $\mu$ l culture media and incubated at 30°C in 5% CO<sub>2</sub> overnight. 100  $\mu$ l 25% formaldehyde was added to each well 35 for 15 minutes. Contents of the wells were removed by inverting the plate. The plate was then washed 3 times with

DPBS. 0.1% phenylhydrazine DPBS solution was added to each well and incubated for 1 hr at 37°C. The plate was inverted and washed 3 times. The plate was blocked with 0.5% BSA-DPBS for 1 hr at room temperature. The plate was inverted and 5 washed 3 times with 1% BPT (PBS containing 1% BSA and 0.05% Tween20). Phage diluted with 1% BPT was added to wells containing fixed cells. Wells without phage added were used to determine background binding of the HRP conjugate. The plates were incubated 2-3 hours on a rotor at room 10 temperature. Plates were washed as before. Plates were incubated with dilute anti-M13-HRP antibody in 1% BPT for 1 hour at room temperature. Following washing, TMB substrate was added and absorbance of the plates were read at 650 nm. Table 10 shows the relative binding of phage encoding 15 peptides to fixed Caco-2 cells.

Table 10.

20		<u>Relative binding of phage encoding peptides to fixed Caco-2 cells</u>	
		<u>Fixed Caco-2 cell binding</u>	
		<u>Phage</u>	
		SNi10	++
		SNi34	+
25		P31	++
		5PAX5	++
		PAX2	+
		HAX42	+
		DCX8	+++
		DCX11	+
		H1	+
30		M13mpl18	-

In vivo phage selection:

35 Further selection of phage expressing peptides capable of binding to the GIT or transporting the GIT was done as follows. The purified library was resuspended in a

buffer, such as TBS or PBS, and introduced onto one side of a tissue barrier, e.g., injected into the duodenum, jejunum, ileum, colon or other *in vivo* animal site using, for instance, a closed loop model or open loop model. Following 5 injection, samples of bodily fluids located across the tissue barrier, e.g., samples of the portal circulation and/or systemic circulation, were withdrawn at predetermined time points, such as 0 to 90 minutes and/or 2 to 6 hours or more. An aliquot of the withdrawn sample (e.g., blood) was used to 10 directly infect a host, e.g., *E. coli*, in order to confirm the presence of phage. The remaining sample was incubated, e.g., overnight incubation with *E. coli* at 37°C with shaking. The amplified phage present in the culture can be sequenced individually to determine the identity of peptides coded by 15 the phage or, if further enrichment is desired, can be precipitated using PEG, and resuspended in PBS. The phage can then be further precipitated using PEG or used directly for administration to another animal using a closed or open GIT loop model system. Portal or systemic blood samples are 20 collected and the phage transported into such circulation systems is subsequently amplified. In this manner, administration of the phage display library with, if desired, repeat administration of the amplified phage to the GIT of the animal, permitted the selection of phage which was 25 transported from the GIT to the portal and/or systemic circulation of the animal.

If desired, following administration of the phage display library to the tissue barrier (e.g., GIT) of the animal model, the corresponding region of the tissue barrier 30 can be recovered at the end of the procedures given above. This recovered tissue can be washed repeatedly in suitable buffers, e.g., PBS containing protease inhibitors and homogenized in, for example, PBS containing protease inhibitors. The homogenate can be used to infect a host, 35 such as *E. coli*, thus permitting amplification of phages which bind tightly to the tissue barrier (e.g., intestinal tissue). Alternatively, the recovered tissue can be

homogenized in suitable PBS buffers, washed repeatedly and the phage present in the final tissue homogenate can be amplified in *E. coli*. This approach permits amplification (and subsequent identification of the associated peptides) of 5 phages which either bind tightly to the tissue barrier (e.g., intestinal tissue) or which are internalized by the cells of the tissue barrier (e.g., epithelial cells of the intestinal tissue). This selection approach of phage which bind to tissues or which are internalized by tissues can be repeated.

10

Treatment of animal tissue barriers  
in vivo with phage display populations

The purified phage display library (random or preselected) was diluted to 500  $\mu$ l in PBS buffer and injected 15 into the closed (or open) intestinal loop model (e.g., rat, rabbit or other species). At time 0 and at successive time points after injection, a sample of either the portal circulation or systemic circulation was withdrawn. An aliquot of the withdrawn blood was incubated with *E. coli*, 20 followed by plating for phage plaques or for transduction units or for colonies where the phage codes for resistance to antibiotics such as tetracycline. The remainder of the withdrawn blood sample (up to 150  $\mu$ l) was incubated with 250  $\mu$ l of *E. coli* and 5 ml of LB medium or other suitable 25 growth medium. The *E. coli* cultures were incubated overnight by incubation at 37°C on a shaking platform. Blood samples taken at other time points (such as 15 min, 30 min, 45 min, 60 min, up to 6 hours) were processed in a similar manner, permitting amplification of phages present in the portal or 30 systemic circulation in *E. coli* at these times. Following amplification, the amplified phage was recovered by PEG precipitation and resuspended in PBS buffer or TBS buffer. The titer of the amplified phage, before and after PEG precipitation, was determined. The amplified, PEG 35 precipitated phage was diluted to a known phage titer (generally between  $10^8$  and  $10^{10}$  phage or plaque forming units (p.f.u.) per ml) and was injected into the GIT of the animal

closed (or open) loop model. Blood samples were collected from portal and/or systemic circulation at various time points and the phage transported into the blood samples were amplified in *E. coli* as given above for the first cycle.

5 Subsequently, the phage was PEG-precipitated, resuspended, titered, diluted and injected into the GIT of the animal closed (or open) loop model. This procedure of phage injection followed by collection of portal and/or systemic blood samples and amplification of phage transported into

10 these blood samples can be repeated, for example, up to 10 times, to permit the selection of phages which are preferentially transported from the GIT into the portal and/or systemic circulation.

15 **6.7. Transport of Phage From Rat Lumen Into the Portal and Systemic Circulation**

Phage from random phage display libraries as well as control phage were injected into the lumen of the rat gastro-intestinal tract (*in situ* rat closed loop model).

20 Blood was collected over time from either the systemic circulation or portal circulation and the number of phage which were transported to the circulation was determined by titering blood samples in *E. coli*.

The phage display libraries used in this study were

25 D38 and DC43 in which gene III codes for random 38-mer and 43-mer peptides, respectively. As a negative control, the identical phage M13mp18, in which gene III does not code for a "random" peptide sequence, was used. Both the library phages D38 and DC43 were prepared from *E. coli*, mixed

30 together, dialyzed against PBS, precipitated using PEG/NaCl and were resuspended in PBS buffer. The M13mp18 control was processed in a similar manner. The titer of each phage sample was determined and the phage samples were diluted in PBS to approximately the same titers prior to injection into

35 the rat closed loop model.

For sampling from the systemic circulation, approximately 15 cm of the duodenum of Wistar rats was tied

off (closed loop model), approximately 0.5ml of phage solution was injected into the closed loop and blood (0.4ml) was sampled from the tail vein at various times. The time points used (in min) were: 0, 15, 30, 45, 60, 90, 120, 180, 5 240 and 300 minutes. For sampling from the portal circulation, the portal vein was catheterized, approximately 15 cm of the duodenum was tied off (closed loop model), 0.5ml of phage solution was injected into the closed loop and blood was sampled from the portal vein catheter at various times. 10 As the portal sampling is delicate, sampling times were restricted to 15, 30, 45 and 60 minutes, where possible. The volume of phage injected into each animal was as follows:

	ANIMALS (15)	VOLUME OF PHAGE INJECTED
15	R1-R3	0.50 ml
	R4	0.43 ml
	R5-R15	0.45 ml

The estimated number of transported phage has been adjusted 20 to account for differences in volume injected into each animal (using 0.5 ml as the standard volume).

To investigate transport into the systemic circulation, animals R1, R2 and R3 received the control phage M13mp18 and animals R4, R5, R6 and R7 received the test phage 25 D38/DC43 mix. To investigate transport into the portal circulation, animals R8, R9 and R10 received the control phage M13mp18 and animals R11, R12, R13 and R14 received the test phage D38/DC43 mix. Animal R15\* received the combined phage samples from animals R4-R7 (see Table 11) which were 30 sampled from the systemic circulation on day one, followed by amplification in *E. coli*, PEG precipitation and resuspension in PBS. On subsequent analysis, the titer of this phage was found to be 100 times greater than the other phage samples used for animals R8-R14. Thus, the data presented for animal 35 R15\* is adjusted down.

Approximately 0.4 ml of the blood was collected at each time point in each model system. 30  $\mu$ l of the collected blood (systemic) was mixed with 100  $\mu$ l of the prepared *E. coli* strain K91Kan, incubated at 37°C for 30 min, and 5 plated out for plaque formation using Top Agarose on LB plates. Various negative controls were included in the titering experiments. The following day, the number of plaque forming units was determined. Similarly, 30  $\mu$ l of the collected blood (portal) and serial dilutions (1:100, 1:1000) 10 thereof was mixed with 100  $\mu$ l of the prepared *E. coli* strain K91Kan, incubated at 37°C for 30 min, and plated out for plaque formation using Top Agarose on LB plates. The following day, the number of plaque forming units was determined.

15 In addition, approximately 300  $\mu$ l of the collected blood from each time point (systemic and portal) was incubated with 5ml of prepared *E. coli* strain K91Kan in modified growth media containing 5mM MgCl<sub>2</sub>/MgSO<sub>4</sub> at 37°C overnight with shaking (to permit phage amplification). The samples were 20 centrifuged and the cell pellet was discarded. Samples of the phage supernatant were collected, serially diluted (10<sup>-2</sup>, 10<sup>-4</sup>, 10<sup>-6</sup>, 10<sup>-8</sup>) in TBS buffer, and plated for plaques in order to determine the number of plaque forming units present in the amplified phage samples.

25 Furthermore, an aliquot of phage was removed from the "amplified" supernatants obtained from test animals R4-R7 (samples from each time point were used), combined, and precipitated using PEG for two hours. The precipitated phage was resuspended in PBS buffer and was injected into closed 30 loop model of animal R15\*, followed by portal sampling.

The number of phage transported from the closed loop model into the systemic circulation is presented in Table 11 hereafter. The number of phage transported from the closed loop model into the portal circulation is presented in 35 Table 12 hereafter. These numbers are corrected for phage input difference and for volume input differences. Clearly, more phage are present in the portal samples than in the

systemic samples, indicative of either hepatic or RES clearance and/or phage instability in the systemic circulation. In addition, the uptake of phage from the GIT into the portal circulation is quite rapid, with substantial 5 number of phages detected within 15 minutes. The results from the portal sampling experiments would also indicate that the kinetics of uptake of phage from the D38/DC43 libraries is quicker than that of the control phage. Thus, there may be preferential uptake of phage coding for random peptide 10 sequences from the GIT into the portal circulation. In the case of animals R13, R14 and R15\*, the % of the phage transported into the titered blood sample within the limited time frame (30, 45 and 15 mins, respectively) was estimated as 0.13%, 1.1% and 0.013%, respectively.

15

TABLE 11

NUMBER OF PHAGE TRANSPORTED FROM THE CLOSED  
LOOP MODEL INTO THE SYSTEMIC CIRCULATION

Time (min)	R1	R2	R3	R4	R5	R6	R7
0	0	0	0	0	0	0	0
15	0	1	9	0	0	1	7
30	2	1	0	0	46	1	11
45	10	4	2	1	32	0	20
60	63	19	21	1	114	0	21
90	104	20	18	3	115	0	22
120	94	24	27	0	64	0	6
180	94	12	23	1	413	0	0
240	14	1	20	0	36	0	0
300	1	1	4	2	0	0	0
Total number of transported phage	382	83	124	8	820	2	87

30

Animals R1, R2 and R3 received the control phage M13mp18.

Animals R4, R5, R6 and R7 received the test phage 35 D38/DC43 mix.

Table 12

NUMBER OF PHAGE TRANSPORTED FROM THE CLOSED  
LOOP MODEL INTO THE PORTAL CIRCULATION

Time (min)	R8	R9	R10	R11	R12	R13	R14	R15*
15	15	6	3	1	19	231,000	1,000,000	20,000
30	1	5	26	-	0	60,000	272,000	-
45	-	1	555	-	1	-	1,240,000	-
60	-	-	-	-	420,000	-	-	-

10 Animals R8, R9 and R10 received the control phage M13mp18.

Animals R11, R12, R13 and R14 received the test phage D38/DC43 mix.

15 Animal R15\* received the combined phage samples from animals R4-R7 (see Table 11) which were sampled from the systemic circulation on day one, followed by PEG precipitation and resuspension in PBS. On subsequent analysis, the titer of this phage was found to be 100 times greater than the other phage samples used for animals R8-R14.

20 Thus, the data measuring phage transport into the portal circulation for animal R15\* is adjusted down.

These studies demonstrated that both the control phage and the D38/DC43 phages are transported over time from the lumen of the GIT into the portal and systemic circulation, as demonstrated by titering the phage transported to the blood in *E. coli*. More phage were transported from the test phage samples into the portal circulation than the corresponding control phage sample. In addition, the kinetics of transport of the test phage into 30 the portal circulation appeared to exceed that of the control phage. Phage from the D38/DC43 libraries which appeared in the systemic circulation of different animals (R4-R7) were pooled, amplified in *E. coli*, precipitated, and re-applied to the lumen of the GIT, followed by collection in the portal circulation and titering in *E. coli*. These selected phage were also transported from the lumen of the GIT into the portal circulation. This *in situ* loop model may represent an

attractive screening model in which to identify peptide sequences which facilitate transport of phage and particles from the GIT into the circulation.

Using this screening model system, a number of 5 preselected phage libraries now exist, including a one pass systemic phage library from animals R4-R7, a one-pass portal library from animals R11-R14, and a two pass, rapid transport, systemic-portal phage library SP-2 from animal R15\*.

10

**6.8. Transport of Phage From Preselected Phage Libraries From the Rat Lumen Into the Portal and Systemic Circulation**

Four preselected phage libraries, GI-D2H, GI-hSI, GI-HPT1 and GI-hPEPT1, were constructed by pooling phage 15 previously selected by screening random phage display libraries D38 and DC43 using the HPT1, HPEPT1, D2H and hSI receptor or binding sites located in the GIT. The phage pools, preselected phage libraries are shown in Table 13. Note that the sequences for PAX2, HAX1, HAX5, HAX6, HAX10, 20 H10 and HAX44 are the same. Also, the sequence for HAX40 is the same as that for H44. The corresponding SEQ ID NOS. are shown in Table 7.

Table 13

25

**PRESELECTED PHAGE LIBRARIES**

	<u>D2H</u>	<u>HSI</u>	<u>HPT1</u>	<u>hPEPT1</u>
	DAB3	S15	HAX9	PAX2 (H10)
	DAB7	S21	HAX35	PAX9
	DAB10	S22	HAX40 (H44)	PAX14
	DAB18	SNi10	HAX42	PAX15
30	DAB24	SNi28	HCA3	PAX16
	DAB30	SNi34	HAX1	PAX17
	DAX15	SNi38	HAX5	PAX18
	DAX23	SNi45	HAX6	PAX35
	DAX24	SNiAX2	HAX10	PAX38
	DAX27	SNiAX6	H40	PAX40
	DCX8	SNiAX8	M13mp18	PAX43
35	DCX11	M13mp18		PAX45
	DCX26			PAX46
	DCX33			P31
	DCX36			P90

DCX39	5PAX3
DCX42	5PAX5
DCX45	5PAX7
M13mp18	5PAX12
	H40
	M13mp18

5

Similar to methods described herein above, these preselected phage libraries together with the negative control phage M13mp18 were injected into the rat closed loop model (6 animals per preselected phage library), blood was collected 10 over time from the portal circulation via the portal vein and, at the termination of the experiment, a systemic blood sample was collected from the tail vein and the intestinal tissue region from the closed loop was collected.

In particular, phages selected *in vitro* to each 15 receptor or binding site located in the GIT were amplified in *E. coli*, PEG-precipitated, resuspended in TBS and the titer of each phage sample was determined by plaquing in *E. coli* as described above. Subsequently, an equal number of each phage (8 x 10<sup>8</sup> phage) for each receptor site was pooled into a 20 preselected phage library together with the negative control phage M13mp18 and each preselected phage library was administered to 6 Wistar rats per library (rats 1-6; GI-D2H, rats 7-12; GI-hSI, rats 13-18; GI-hPEPT1, and rats 19-24; GI-HPT1). Using the *in situ* loop model described above, 0.5 ml 25 of preselected phage library solution was injected into the tied-off portion of the duodenum/jejunum. Blood was collected into heparinized tubes from the portal vein at 0, 15, 30, 45 and 60 minutes. A blood sample was taken from the systemic circulation at the end of the experiment. 30 Similarly, the portion of the duodenum/jejunum used for phage injection was taken at the end of the experiment.

Thirty microliters of the collected portal blood (neat and 10<sup>-2</sup>, 10<sup>-4</sup>, 10<sup>-6</sup> dilutions) was added to 30 µl *E. coli* K91Kan cells (overnight culture) and incubated at 37°C for 10 35 min. Subsequently, 3 ml of top agarose was added and the samples were plated for plaques. One hundred microliters of

the collected portal blood was added to 100 $\mu$ l of *E. coli* K91Kan. Five milliliters of LB medium was then added and the samples were incubated at 37°C overnight in a rotating microbial incubator. The *E. coli* was removed by 5 centrifugation and the amplified phage supernatant samples were either titered directly or were PEG-precipitated, resuspended in TBS and titered. Following titration of the amplified phage, samples containing phage from each set of animals were combined, adjusting the titer of each sample to 10 the same titer, and were plated for plaques on LB agar plates (22cm<sup>2</sup> square plates). Either 12,000 or 24,000 phage were plated for plaques.

Thirty microliters of the collected systemic blood (neat and 10<sup>-2</sup>, 10<sup>-4</sup>, 10<sup>-6</sup> dilutions) was added to *E. coli* 15 K91Kan cells, incubated at 37°C for 10 min. Three ml of top agarose was then added and the samples were plated for plaques. One hundred microliters of the collected systemic blood was added to 100 $\mu$ l of *E. coli* K91Kan, incubated at 37°C for 10 min. Five milliliters of LB medium was then added and 20 the samples were incubated at 37°C overnight in a rotating microbial incubator. The *E. coli* was removed by centrifugation and the amplified phage supernatant samples were either titered directly or were PEG-precipitated, resuspended in TBS and titered. Following titration of the 25 amplified phage, samples containing phage from each set of animals were combined, adjusting the titer of each sample to the same titer, and were plated for plaques on LB agar plates (22cm<sup>2</sup> square plates). Either 12,000 or 24,000 phage were plated for plaques.

30 The intestinal tissue portion used in each closed loop was excised. The tissue was cut into small segments, followed by 3 washings in sterile PBS containing protease inhibitors, and homogenized in an Ultra thorex homogeniser (Int-D samples). Alternatively, the tissue (in PBS 35 supplemented with protease inhibitors) was homogenized in an Ultra Thorex homogenizer, washed 3 times in PBS containing protease inhibitors and resuspended in PBS containing

protease inhibitors (Int-G samples). In each case, serial dilutions (neat and 10<sup>-2</sup>, 10<sup>-4</sup>, 10<sup>-6</sup> dilutions) of the tissue homogenate was titered in *E. coli*. In addition, an aliquot (100 $\mu$ l) of the tissue homogenate was added to 100 $\mu$ l of 5 *E. coli* K91Kan, incubated at 37°C for 10 min, followed by addition of 5ml of LB medium and incubation overnight at 37°C in a rotating microbial incubator.

The phage amplified from the portal blood, systemic blood and intestinal tissue was plated for plaques. The 10 plaques were transferred to Hybond-N Nylon filters, followed by denaturation (1.5M NaCl, 0.5M NaOH), neutralization (0.5M TRIS-HCl, pH7.4, 1.5M NaCl), and washing in 2X SSC buffer. The filters were air-dried, and the DNA was cross-linked to the filter (UV crosslinking: 2min, high setting). The 15 filters were incubated in pre-hybridization buffer (6X SSC, 5X Denhardt's solution, 0.1% SDS, 20 $\mu$ g/ml yeast tRNA) at 40°C-45°C for at least 60 min.

Synthetic oligonucleotides, (22-mers), complimentary to regions coding for the receptor or binding 20 sites used to create the preselected phage library, were synthesized (see Table 14 below).

Table 14

OLIGONUCLEOTIDES USED IN IN VIVO SCREEN

25	CLONE NAME	OLIGO	SEQ. ID. NO.
	S15	5' TCCGGACTCTCATAAGCGCCGG'	111
	S21	5' ACAACGGGCCAGAAAGAGCGAG'	112
	S22	5' ACACCACCCCAATCGGAGCTAC'	113
30	SNi10	5' TCAGAATCCGTGCACTGGCAA'	114
	SNi28	5' GCCCTATTCATACCACCGGAGT'	115
	SNi34	5' CATCAGTCCTACCGCCGAAAAG'	116
	SNi38	5' CGTATAGCTATTGTGAGCGATG'	117
	SNi45	5' ACGCGCGGAACGAGCAGTACCA'	118
	SNiAX2	5' CCATAATGATCCCCGTCACTAT'	119
35	SNiAX6	5' AGACACCCCTTAGCCGTCGTAG'	120
	SNiAX8	5' AGCTCCGTGACCTTAGTCATAA'	121

CLONE NAME	OLIGO	SEQ. ID. NO.
DAB3	5' TGCACAGCTCAGCGCCGCACCA 3'	122
DAB7	5' ACGGGTCATCAGCGCCGCACCA 3'	123
DAB10	5' TGTACACCCCTCCCCGGACTT 3'	124
5 DAB18	5' ACTCGCAATTATTGGCGCTCGA 3'	125
DAB24	5' GTCTTCTCAACCTTATCCTGCG 3'	126
DAB30	5' AAAGCCCCCTGCTAAACTCCCA 3'	127
DAX15	5' CTGCGTCTGCCACGTCGTCATC 3'	128
DAX23	5' GTTAAAAGAGGGCAAGCTCGGA 3'	129
10 DAX24	5' CCGAGTTCTTGATGTCCTCCAT 3'	130
DAX27	5' TCCAATGCCTGTACCACGGATG 3'	131
DCX8	5' TCGCAACCGATATCGTGCTTAT 3'	132
DCX11	5' TGCATACACTGCTTGGAGCCCT 3'	133
DCX26	5' GAAATCTCACTAGTAGTCCGCC 3'	134
15 DCX33	5' GCGGGCAAGACAGTCCAATTCC 3'	135
DCX36	5' GAGCTCCAATTCCACGACGACC 3'	136
DCX39	5' GGTTGCCATGCGTTCAAACCTAC 3'	137
DCX42	5' TCCCGGGGGACAAACCCGAAT 3'	138
DCX45	5' CTGCTAGTCTTATCATTCCCCA 3'	139
20 PAX2	5' CTATCGACACTATAGGGCCTAC 3'	140
PAX9	5' TACCCCTTGTAAACCCACACTAGG 3'	141
PAX14	5' TTCTTCTGAATAGACCGGCCGA 3'	142
PAX15	5' CCACCACCCCTAACCCGACAAT 3'	143
PAX16	5' AGGGGGAGACTTGTTCACAAAC 3'	144
25 PAX17	5' CGGCTCATACCAACCGAAAGCTA 3'	145
PAX18	5' ATCGTCCTACTGTAATCCTCGA 3'	146
PAX35	5' GACACACTACTCAGGTCCACCT 3'	147
PAX38	5' CCATAATCAACATTGCCGCCCT 3'	148
PAX40	5' CAAAACGCTGCCCAAACCTCA 3'	149
30 PAX43	5' GTAAACTTGTGCTTCTCGCACC 3'	150
PAX45	5' CCATGGTCCGGGTACACCTGAA 3'	151
PAX46	5' GTTACTAACGGAACGAGACCTA 3'	152
P31	5' TGTGGCGTTCTCAACCCGTT 3'	153
P90	5' ACAACCGGAGTTGTTCTGCCTA 3'	154
35 5PAX3	5' TAAGCATCGGCCACGTTCTTCG 3'	155
5PAX5	5' TTATCCCTGGTGTGCAGGTTGA 3'	156

CLONE NAME	OLIGO	SEQ. ID. NO.
5PAX7	5' TATCAGAATCGTAGTCGGACGG <sup>3'</sup>	157
5PAX12	5' CTTTGTAAATGGAACCACAACCC <sup>3'</sup>	158
HAX9	5' CGGTGGCTCATCTCCCTCTTAT <sup>3'</sup>	159
5 HAX35	5' ATCAGACTGGCTGGGACCAACAA <sup>3'</sup>	160
HAX40	5' CACAAACCTCCTCTCCGCGAACT <sup>3'</sup>	161
HAX42	5' AGATTCTGTCCCCAACGCGTGAT <sup>3'</sup>	162
HCA3	5' GGGAAATTGCAAAGCTATACTC <sup>3'</sup>	163
H40	5' CCCCCGTGGAATTCAACCTGTGA <sup>3'</sup>	164
10 M13 (positive)	5' GTCGTCTTCCAGACGT <sup>3'</sup>	165
M13 (negative)	5' CTTGCATGCCTGCAAGTCGAC <sup>3'</sup>	166

The oligonucleotides (5pmol) were 5'end labelled with <sup>32</sup>P-ATP and T4 polynucleotide kinase and approximately 2.5pmol of 15 labelled oligonucleotide was used in hybridization studies. Hybridizations were performed at 40-45°C overnight in buffer containing 6X SSC, 5X Denhardt's solution, 0.1% SDS, 20µg/ml yeast tRNA and the radiolabeled synthetic oligonucleotide, followed by washings (20-30 min at 40-45°C) in the following 20 buffers: (i) 2X SSC / 0.1% SDS, (ii) 1X SSC / 0.1% SDS, (iii) 0.1X SSC / 0.1% SDS. The filters were air-dried and exposed for autoradiography for 15 hours, 24 hours or 72 hours.

Hybridization data indicated that all the 25 oligonucleotide probes bound specifically to their phage target except for the HAX9 probe which apparently was not labeled. A negative control probe that hybridized only to M13mp18 DNA showed a weak to negative signal in all samples tested (data not shown).

Hybridization data for pools from each receptor 30 group of rats was compiled. Tables 15, 16, 17 and 18 show a representative compilation of autoradiograph signals of the HSI, D2H, HPT1 and hPEPT1 receptor groups. These Tables show the phage absorption and uptake from the closed loop GIT model to portal and systemic circulation and phage 35 absorption/internalization to intestinal tissue. In these Tables, Int-G refers to intestinal tissue homogenized prior

to washing and recovery while Int.-D refers to intestinal tissue washed prior to homogenization and phage recovery. In all cases, leading phage candidates were present in more than one animal.

5

Table 15

## SUMMARY OF AUTORADIOGRAPH SIGNALS OF HSI ANIMAL STUDY

10

15

20

Phage	Portal	Int.-G	Int.-D
S15	++	+/-	+/-
S21	-	-	-
S22	-	-/+	-
SNi-10	+++/+	++	++
SNi-28	-	-	-
SNi-34	++	-	-
SNi-38	++	-	-
SNi-45	-	-	-
SNiAX-2	-	-	-
SNiAX-6	-	-	-
SNiAX-8	-	-	-
M13	+++++	+++++	+++++
M13	nd*	+	-

\*not detected

25

30

35

Table 16

## SUMMARY OF AUTORADIOGRAPH SIGNALS OF D2H ANIMAL STUDY

	Phage	Portal	Int.-G	Int.-D
5	DAB3	+++	+/-	-/+
	DAB7	++	++	-/+
	DAB10	++++++	+/-	-/+
	DAB18	-	-	-
	DAB24	-	-	-
	DAB30	++++	++	+++
	DAX15	-	-	-
	DAX23	-/+	+	-/+
	DAX24	-	-	-
	DAX27	-	+	-
10	DCX8	++++	+/-	-
	DCX11	+++++	++	-/+
	DCX26	-	-	-
	DCX33	+++	++	++
	DCX36	-	-	-
	DCX39	-	-/+	-
	DCX42	-	-	-/+
	DCX45	-	++	-
	M13 (+)	+++++	+++++	+++++
	M13 (-)	+/-	-/+	-

20

Table 17

## SUMMARY OF AUTORADIOGRAPH SIGNALS OF HPT1 ANIMAL STUDY

	Phage	Int.-G	Portal	Systemic
25	H40	-	-	++++
	HAX9	ND	ND	ND
	HAX35	-	+	-
	HAX40	-	-	-
	HAX42	-	++	++
	HCA3	-	-	-
	PAX2	-	+++	++++
	M13 (+)	++++++	++++++	++++++
	M13 (-)	-	--/+	-

35

Table 18

SUMMARY OF AUTORADIOGRAPH SIGNALS OF hPEPT1 ANIMAL STUDY

	Phage	Int.-G	Portal	Systemic
5	PAX2	-	++	-
	PAX9	++	+++	-
	PAX14	-	++	-
	PAX15	-/+	-	-
	PAX16	-	-	-
	PAX17	+	++/+	-
	PAX18	-	-	-
	PAX35	-	-	-
	PAX38	-/+	-	-
	PAX40	+	+++	-
10	PAX43	+	-	-
	PAX45	-	-	-
	PAX46	-	+++	-
	P31	++	++++	++
	5PAX3	++/+	++	-
	5PAX5	-	-	++
	5PAX7	+++	-	-
	5PAX12	++++	++	-
	H40	++	++	-
	M13 (+)	++++++	++++++	++++++
15	M13 (-)	-	-	-

20

Apart from the synthetic oligonucleotide to HAX9, all oligonucleotides were initially confirmed to be radiolabeled, as determined by hybridization to the corresponding phage target (eg., phage S15 hybridized to the oligonucleotide S15). In addition, under the experimental conditions used, the oligonucleotides essentially did not hybridize to the negative control phage template M13mp18. Two oligonucleotides were synthesized to the phage M13mp18: (1) a positive oligonucleotide which hybridizes to a conserved sequence in both M13mp18 and each of the GIT receptor or GIT binding site selected phages [designated M13 (positive)]; and (2) a negative oligonucleotide which only hybridizes to a sequence unique to the multiple cloning site of phage M13mp18 and which does not hybridize to any of the GIT receptor or GIT binding site selected phages.

In the case of the hSI pool of phages, only four phages were transported from the closed loop model into the portal circulation: phages S15, SNi-10, SNi-34 and SNi-38. The other phages, S21, S22, SNi-28, SNi-45, SNiAX-2, SNiAX-6 and 5 SNiAX-8, were not transported from the GIT into the portal circulation. In addition, phages SNi-10 and to a lesser extent phages S15 and S22 were found in the intestine samples or fractions, whereas the other phages were not. There was a very low presence (<0.1%) of the phage M13mp18 in the Int-G 10 samples. These results show that phages can be further selected from pre-selected libraries, permitting the identification of phages which are transported from the GIT closed loop into the portal circulation or phages which bind to or are internalized by intestinal tissue.

15 In the case of the D2H pool of phages, there was a rank order by which phages were transported from the GIT closed loop model into the portal circulation, with phages DCX11 and DAB10 preferably transported, followed by phages DCX8, DAB30, DAB3 and DAB7. A number of phages from this pool were not 20 transported into the portal circulation, including phages DAB18, DAB24, DAX15, DAX24, DAX27, DCX26, DCX36, DCX39, DCX42, DCX45. There is a very low level of transport of phage DAX23 from the GIT into the portal circulation. Similarly, only some of the phages were found in the intestinal samples 25 fractions, including phages DAB30, DCX33, DAB7, DCX11, DCX45 and to a much lesser extent phages DAB3, DAB10, DCX8, DCX39, DCX42. Some phages were not found in the intestinal samples, including phages DAB18, DAB24, DAX15, DAX24, DCX26, and DCX36. There was a very low presence (<0.1%) of the phage 30 M13mp18 in the Int-G samples. These results showed that phages can be further selected from pre-selected libraries, permitting the identification of phages which are transported from the GIT closed loop into the portal circulation or phages which bind to or are internalized by intestinal 35 tissue.

In the case of the HPT1 pool of phages, there was a rank order by which phages were transported from the GIT closed

loop model into the portal or systemic circulation. Phage PAX2 (which was used at a 4X concentration relative to the other phages in this pool) followed by phage HAX42 was found in the portal and systemic circulation; phage H40 was found 5 in the systemic circulation only. None of the phages in this pool were found in the intestine samples or fractions. Phage M13mp18 was not found in the intestine fractions or systemic circulation, with very low incidence (<0.001%) in the portal circulation. These results show that phages can be further 10 selected from pre-selected libraries, permitting the identification of phages which are transported from the GIT closed loop into the portal and/or systemic circulation or phages which bind to or are internalized by intestinal tissue.

15 In the case of the hPEPT1 pool of phages, the phages PAX2 and H40 were also included in this pool. A number of phages from this pool were found in the portal circulation, including phages P31 (SEQ ID NO:43), PAX46, PAX9, H40, PAX17, PAX40, PAX2, PAX14, 5PAX3 and 5PAX12. A number of phages 20 were not found in the portal blood including the negative control phage M13mp18, PAX15, PAX16, PAX18, PAX35, PAX38, PAX43, PAX45, P90, 5PAX5 and 5PAX7. The only phage found in the systemic circulation were phages 5PAX5 and P31 (SEQ ID NO:43). In addition, there was preferential binding of some 25 phages to the intestine, including phages 5PAX12, 5PAX7, 5PAX3, H40, P31 (SEQ ID NO:43), PAX9, and to a lesser extent phages PAX38 and PAX15. Some phages were not found in the intestine samples, including the negative control phage M13mp18 and the phages PAX2, PAX14, PAX16, PAX18, PAX35, 30 PAX45, PAX46, P90 and 5PAX5. These results show that phages can be further selected from pre-selected libraries, permitting the identification of phages which are transported from the GIT closed loop into the portal and/or systemic circulation or phages which bind to or are internalized by 35 intestinal tissue.

Further Characterization of Select Sequences

Following initial screening of the four recombinant receptor sites (hPEPT1, HPT1, D2H, hSI) of the gastrointestinal tissue, with the phage display libraries, a 5 series of phage were isolated which showed preferential binding to the respective target receptor sites in comparison to negative control protein BSA protein and the recombinant protein recombinant human tissue factor (hTF) (which, like the recombinant receptors of the gastrointestinal tissue, 10 contained a poly-histidine tag at its NH<sub>2</sub>-terminal end). In subsequent experiments same titers of the selected phage which bound to each target receptor site were combined into a single pool (i.e., one pool of HPT1 binding phage, one pool of hPEPT1 binding phage, one pool of D2H binding phage, and 15 one pool of hSI binding phage). Each pool was supplemented with an equivalent titer of the negative control phage M13mp18. These phage pools were injected into a closed duodenal loop region of rat intestinal tissue and subsequently phage was harvested and recovered which was 20 bound to and retained by the intestinal tissue and/or was absorbed from the intestinal loop into the portal and/or systemic circulation. In addition, a selection of the initial phages which bound to the target recombinant receptor site were analyzed for binding to either fixed Caco-2 cells 25 and/or to fixed C2BBe1 cells. The selection of the final lead peptide sequences was based on the ability of the phage, coding for that peptide sequence (1) to bind to the target recombinant receptor site *in vitro* in preference to its binding to the negative control proteins BSA and/or hTFs, (2) 30 to bind to rat intestinal tissue following injection into a closed duodenal loop of rat intestinal tissue in preference to the negative control phage M13mp18, (3) to be absorbed from rat intestinal tissue into either the portal and/or systemic circulation following injection into a closed 35 duodenal loop of rat intestinal tissue in preference to the negative control phage M13mp18, and (4) to bind to either fixed Caco-2 cells or fixed C2BBe1 cells in phage binding

studies in preference to the negative control phage M13mp18. Peptides were also selected with consideration to the ease of chemical synthesis.

5        6.9. GST Fusion Proteins of GIT Targeting Peptides  
          Construction of GST Fusion Proteins of GI  
          Targeting Peptides

Glutathione S-transferase (GST) vectors encoding fusion proteins of GI targeting peptides were constructed in the vector pGEX4T-2 (source, Pharmacia Biotech, Piscataway, 10 NJ). Briefly, single-strand DNA from the clones of interest were amplified by the polymerase chain reaction. The amplified DNA was then cleaved with the restriction enzymes KhoI and NotI and then ligated into SalI/NotI cleaved pGEX4T-2. Following transformation, the DNA sequence for 15 each construct was verified by sequencing.

For construction of the truncated versions of the GST fusion proteins, where the inserted sequence was less than 45 base pairs, overlapping oligonucleotides containing cohesive SalI and NotI termini, and encoding the sequence of 20 interest, were annealed and then ligated directly into SalI/NotI cleaved pGEX4T-2. Following transformation, the DNA sequence for each construct was verified.

A diagrammatic representation of the various GST fusion protein constructs that have been synthesized is 25 indicated in Figures 5A-5C.

Expression and Purification of GST Fusion Proteins

*Escherichia coli* BL21 cells containing GST fusion protein constructs were grown overnight in 2X YT media 30 containing 100 µg/ml ampicillin (2X YT/amp). Overnight cultures were diluted 1:100 in 2X YT broth (100 ml), and cells were grown to an  $A_{600}$  of 0.5 at 30°C, induced with 1mM isopropyl-1-thio-B-D-galactopyranoside, and grown for an additional 3 h. Cells were harvested by centrifugation and 35 resuspended in 5 ml of PBS containing a mixture of the proteinase inhibitors (Boehringer/Mannheim). Cells were

sonicated on ice, and the cell lysates were centrifuged at 12,000 x g for 10 minutes at 4°C. Supernatant fractions were reacted for 30 minutes at room temperature with 2 ml of a 50% slurry of glutathione-Sepharose® 4B, washed 3 times with 1.5 5 ml of PBS (at room temperature), and the bound GST fusion proteins were eluted by reaction for 10 minutes at room temperature with 3 X 1ml of 10 mM reduced glutathionein 50 mM Tris HCl pH 8.0. Protein was quantified by the Bio-Rad protein assay followed by characterization by SDS-10 polyacrylamide gel electrophoresis.

#### ELISA of GST fusion peptides

The standard ELISA procedure was modified as follows. GST proteins were diluted to an appropriate 15 concentration in PBS containing 1%BSA and 0.05% Tween20 (1%BPT), titered and incubated one hour at room temperature. Following five washes an anti-GST monoclonal antibody was added (Sigma, St. Louis Clone GST-2 diluted 1:10,000 in 1%BPT) and incubated one hour. After five more washes goat 20 anti-mouse IgG2b-HRP was added (Southern Biotechnology Associates Inc., Birmingham, AL, diluted 1:4000 in 1%BPT) and incubated one hour. After five washes plates were developed with TMB peroxidase substrate (Kirkegard and Perry, Gaithersburg, MD). All data is presented with background 25 binding subtracted.

Figure 6 shows the binding of GST-SNI10, GST-SNI34 and GST alone to the hSI receptor and to fixed C2BBe1 cells.

#### GST Fusion Proteins of Selected GIT Targeting Peptides

30 Results show that GST-DXB8, GST-PAX2, GST-P31, GST-SNI10 and GST-SNI34 bound fixed Caco-2 or C2BBe1 cells (Figures 7 and 8) relative to GST control binding. GST-HAX42, GST-5PAX5, all showed weak to moderate binding relative to GST control.

35 Interestingly, P31 truncation 103-GST fusion protein bound almost as well as full-length P31 (SEQ ID NO:43) to fixed Caco-2 cells (A). This suggests the portion

of the P31 sequence (SEQ ID NO:43) responsible for binding resides in this portion. PAX2.107 bound similarly to full-length PAX2; therefore, this portion most likely contains the amino acid sequence responsible for binding (B). In 5 preliminary assays, none of the DCX8 truncations bound similarly to full-length DCX8 to Caco-2 cells suggesting the binding region spans more than one of these pieces.

Inhibition of Binding by Synthetic Peptides

10 Binding of GST-P31 to fixed C2BBe1 Cells

The standard ELISA procedure was modified as follows. GST fusion proteins and peptides were diluted to an appropriate concentration in PBS containing 1% BSA and 0.05% Tween 20. Peptides were titered, a constant concentration of 15 diluted GST protein was added to titered peptides and the mixture was incubated one hour at room temperature. Following five washes, an anti-GST monoclonal antibody was added (Sigma, St. Louis Clone GST-2 diluted 1:10,000 in 1% BPT) and incubated one hour. After five more washes goat 20 anti-mouse IgG2b-HRP was added (Southern Biotechnology Associates Inc., Birmingham, AL, diluted 1:4000 in 1% BPT) and incubated one hour. After five washes plates were developed with TMB peroxidase substrate (Kirkegard and Perry, Gaithersburg, MD). All data is presented with background 25 binding subtracted.

Figures 9A and 9B show the inhibition of GST-P31 binding to C2BBe1 fixed cells. The peptide competitors are ZElan024 which is the dansylated peptide version of P31 (SEQ ID NO:43) and ZElan044, ZElan049 and ZElan050 which are 30 truncated, dansylated pieces of P31 (SEQ ID NO:43). Data is presented as O.D. vs. peptide concentration and as percent inhibition of GST-P31 binding vs. peptide concentration. Uncompeted GST-P31 binding was considered as 100% binding. IC<sub>50</sub> values are estimates using the 50% line on the percent 35 inhibition graph.

GST-P31 and GST-PAX2 exhibited no crossreactive binding to ZElan024 (P31) (SEQ ID NO:43) and ZElan018 (PAX2)

at the 0.5  $\mu$ g/ml concentration used in competition assays. GST-HAX42 exhibited crossreactivity to ZElan018 (PAX2) and ZElan021 (HAX42) at the 5  $\mu$ g/ml concentration used in competition assays.

5 Figures 10A-10C present a compilation of data generated by competition ELISA of GST-P31, GST-PAX2, GST-SNi10 and GST-HAX42 versus various dansylated peptides on fixed C2BBe1 cells.  $IC_{50}$  values are in  $\mu$ M and include ranges determined from multiple assays. The GST/C2BBe1 column is a  
10 summary of GST protein binding to fixed C2BBe1 cells.

#### Binding to fixed Caco-2 Cells

Caco-2 cells were fixed, treated with phenylhydrazine and blocked as described above. Synthetic 15 peptides (100 $\mu$ g/ml) were applied in duplicate to Caco-2 cells and serially diluted down the 96-well plate. The corresponding GST-peptide fusion protein (10 $\mu$ g) was added to each well and the plates were incubated for 2h at room temperature with agitation. Binding of the GST-peptide 20 fusion proteins to the cells was assayed using the ELISA technique described above. GST-P31 binding was inhibited by ZElan024, ZElan028 and ZElan031 as well as the two D forms ZElan053 and ZElan054. GST-PAX2 binding was inhibited by ZElan032, ZElan033, and ZElan035. GST-HAX42 binding was not 25 inhibited by ZElan021 (full length HAX42) but it was inhibited by ZElan018 (PAX2) and ZElan026 and ZElan038 (scrambled PAX2 peptides).

#### Transport and Uptake of GST-Peptide Fusions into Live Caco-2 Cells

30 Transport and uptake of GST-peptide fusions and deletion derivatives across cultured polarized Caco-2 monolayers over 4 hours in HBSS buffer was examined using an anti-GST ELISA assay. In another experiment, transport and 35 uptake of GST-peptide fusions and deletion derivatives across

cultured polarized Caco-2 monolayers over 24 hours in serum-free medium (SFM) was examined using an anti-GST ELISA assay.

#### Materials

5            Buffered Hank's balanced salt solution (bHBSS) = 1x HBSS (Gibco CN.14065-031) supplemented with 0.011M glucose (1g/l), 25 mM Hepes (15 mM acid (3.575g/l; Sigma CN.H3375); 10mM base (2.603g/l; Sigma CN.H1016)].

Chloroquine: Made up as 10mM solution in water  
10 [Sigma CN C6628]

Lysate buffer: 30 mM Tris-HCl pH8.0; 1mM EDTA

Serum-free medium (SFM) is normal medium without serum.

#### 15 Method

a)    4h HBSS study: Transepithelial electrical flux (TER) across the Caco-2 monolayers grown on snapwells (passage 33; 23 days old) was measured to confirm monolayer integrity before beginning the experiment. The medium was 20 removed and the cells were washed once with bHBSS. bHBSS containing 100 $\mu$ M chloroquine was added and the cells were incubated for 2h at 37°C. The bHBSS+chloroquine was replaced with 0.5ml bHBSS containing GST-peptide fusions (100 $\mu$ g/ml) and the cells were incubated as before. Basolateral samples 25 were removed at the following times: 0, 0.5h, 2h, and 4h.

At 4h, TER was measured, the apical medium was sampled and the apical reservoir was washed 6 times with HBSS. The cells were allowed to lyse for 1h on ice in lysate buffer, after which, lysate sample was collected. All samples were stored 30 at -70°C until assay by anti-GST ELISA. Before analysis, samples were normalized for protein content relative to each other using a BioRad protein assay.

b)    24h SFM study: Transepithelial electrical flux (TER) across the Caco-2 monolayers grown on snapwells 35 (passage 33; 23 days old) was measured to confirm monolayer integrity before beginning the experiment. The medium was removed and the cells were washed once with SFM. SFM

containing GST-peptide fusions (100 $\mu$ g/ml) was added to the cells which were incubated at 37°C for 24h at 5% CO<sub>2</sub>. After 24 hours, TER readings were taken, and samples from the basolateral and apical reservoirs were removed. The apical 5 reservoir was washed 6 times with PBS. The cells were allowed to lyse for 1h on ice in lysate buffer, after which lysate sample was collected. All samples were stored at -70° until assay by anti-GST ELISA. Before analysis, samples were normalized for protein content relative to each other using a 10 BioRad protein assay.

### Results

All of the GST-peptide fusions and controls examined were transported across live Caco-2 monolayers. 15 Full-length GST-P31 and GST-DCX8, but not truncations of these molecules had a higher flux than GST alone.

Internalization of GST-peptide fusions into polarized Caco-2 cells was investigated in two experiments. In experiment 1, 15 $\mu$ g of GST-peptide fusion was applied in 20 bHBSS and internalized GST-peptide was recovered by lysing the cells after 4h. In experiment 2, 10 $\mu$ g of GST-peptide was applied in either a) bHBSS (lysate recovered after 4h), or b) serum-free medium (lysate recovered after 24h).

Figure 11A describes complete transport of GST-peptide across a polarized Caco-2 monolayer and does not necessarily refer to internalization, i.e., the GST-peptide was recovered from the basolateral reservoir of a snapwell but the proteins could have crossed the barrier by the paracellular route.

30

### Effect of Thrombin Cleavage on Binding of GST-Peptide Fusions to Fixed Caco-2 Cells

Binding of intact and thrombin-cleaved GST-peptide fusions to fixed Caco-2 cells was compared. Reduced binding 35 of the thrombin-cleaved GST-peptide fusions relative to intact fusions indicates that the peptide component of the fusion, and not the GST domain, mediates binding.

Method

Confluent Caco-2 monolayers grown in 96-well plates (p38) were fixed and treated with 0.1% phenylhydrazine before blocking with 0.1% BSA in PBS. Thirty micrograms of each 5 GST-peptide was treated with bovine thrombin (1 $\mu$ /ml; 0.4 NIH units; Sigma CN.T9681) for 18h at room temperature in 20mM Tris-HCl pH8.0, 150mM NaCl, 2.5mM CaCl<sub>2</sub>. Controls were similarly treated without addition of thrombin. Ten 10 micrograms of each GST-peptide fusion was removed for PAGE analysis, and 10 $\mu$ g of fusions were added in duplicate to the fixed Caco-2 cells before 5-fold serial dilutions (1% BPT diluent). The fusions were allowed to bind for 1h at room temperature. Following 6 washes with 1% BPT, binding was assayed by ELISA.

15

Results

Results are shown in Figure 12.

Conclusions:

20 PAGE analysis confirmed that the GST-peptide fusions were effectively cleaved with thrombin. Cleavage with thrombin significantly reduced detection of binding of GST-P31.103, GST-PAX2.106, GST-DCX8, GST-SNi10 to fixed Caco-2 cells, indicating that the peptide component, and not the 25 GST domain, mediates binding.

### 6.10. Synthesis of Peptides

#### 6.10.1. Procedure For Solid Phase Synthesis

Peptides may be prepared by methods that are known 30 in the art. For example, in brief, solid phase peptide synthesis consists of coupling the carboxyl group of the C-terminal amino acid to a resin and successively adding N-alpha protected amino acids. The protecting groups may be any known in the art. Before each new amino acid is added to 35 the growing chain, the protecting group of the previous amino acid added to the chain is removed. The coupling of amino acids to appropriate resins is described by Rivier et al.,

U.S. Patent No. 4,244,946. Such solid phase syntheses have been described, for example, by Merrifield, 1964, J. Am. Chem. Soc. 85:2149; Vale et al., 1981, Science 213:1394-1397; Marki et al., 1981, J. Am. Chem. Soc. 103:3178 and in U.S. 5 Patent Nos. 4,305,872 and 4,316,891. In a preferred aspect, an automated peptide synthesizer is employed.

By way of example but not limitation, peptides can be synthesized on an Applied Biosystems Inc. ("ABI") model 431A automated peptide synthesizer using the "Fastmoc" 10 synthesis protocol supplied by ABI, which uses 2-(1H-Benzotriazol-1-yl)-1,1,3,3,-tetramethyluronium hexafluorophosphate ("HBTU") (R. Knorr et al., 1989, Tet. Lett., 30:1927) as coupling agent. Syntheses can be carried out on 0.25 mmol of commercially available 15 4-(2',4'-dimethoxyphenyl-(9-fluorenyl-methoxycarbonyl)-aminomethyl)-phenoxy polystyrene resin ("Rink resin" from Advanced ChemTech) (H. Rink, 1987, Tet. Lett. 28:3787). Fmoc amino acids (1 mmol) are coupled according to the Fastmoc protocol. The following side chain 20 protected Fmoc amino acid derivatives are used:  
FmocArg(Pmc)OH; FmocAsn(Mbh)OH; FmocAsp(<sup>t</sup>Bu)OH;  
FmocCys(Acm)OH; FmocGlu(<sup>t</sup>Bu)OH; FmocGln(Mbh)OH; FmocHis(Tr)OH;  
FmocLys(Boc)OH; FmocSer(<sup>t</sup>Bu)OH; FmocThr(<sup>t</sup>Bu)OH;  
FmocTyr(<sup>t</sup>Bu)OH. [Abbreviations: Acm, acetamidomethyl; Boc, 25 tert-butoxycarbonyl; <sup>t</sup>Bu, tert-butyl; Fmoc, 9-fluorenylmethoxycarbonyl; Mbh, 4,4'-dimethoxybenzhydryl; Pmc, 2,2,5,7,8-pentamethylchroman-6-sulfonyl; Tr, trityl].

Synthesis is carried out using N-methylpyrrolidone (NMP) as solvent, with HBTU dissolved in 30 N,N-dimethylformamide (DMF). Deprotection of the Fmoc group is effected using approximately 20% piperidine in NMP. At the end of each synthesis the amount of peptide present is assayed by ultraviolet spectroscopy. A sample of dry peptide resin (about 3-10 mg) is weighed, then 20% piperidine in DMA 35 (10 ml) is added. After 30 min sonication, the UV (ultraviolet) absorbance of the dibenzofulvene-piperidine adduct (formed by cleavage of the N-terminal Fmoc group) is

recorded at 301 nm. Peptide substitution (in mmol g<sup>-1</sup>) can be calculated according to the equation:

$$\text{substitution} = \frac{A \times v}{7800 \times w} \times 1000$$

5

where A is the absorbance at 301 nm, v is the volume of 20% piperidine in DMA (in ml), 7800 is the extinction coefficient (in mol<sup>-1</sup>dm<sup>3</sup>cm<sup>-1</sup>) of the dibenzofulvene-piperidine adduct, and w is the weight of the peptide-resin sample (in mg).

10 Finally, the N-terminal Fmoc group is cleaved using 20% piperidine in DMA, then acetylated using acetic anhydride and pyridine in DMA. The peptide resin is thoroughly washed with DMA, CH<sub>2</sub>Cl<sub>2</sub>, and finally diethyl ether.

15

#### 6.10.2. Cleavage and Deprotection

By way of example but not limitation, cleavage and deprotection can be carried out as follows: The air-dried peptide resin is treated with ethylmethyl-sulfide (EtSMe), ethanedithiol (EDT), and thioanisole (PhSMe) for approximately 20 min. prior to addition of 95% aqueous trifluoracetic acid (TFA). A total volume of approximately 50 ml of these reagents are used per gram of peptide-resin. The following ratio is used: TFA:EtSMe:EDT:PhSMe (10:0.5:0.5:0.5). The mixture is stirred for 3 h at room temperature under an atmosphere of N<sub>2</sub>. The mixture is filtered and the resin washed with TFA (2 x 3 ml). The combined filtrate is evaporated *in vacuo*, and anhydrous diethyl ether added to the yellow/orange residue. The resulting white precipitate is isolated by filtration. See King et al., 1990, Int. J. Peptide Protein Res. 36:255-266 regarding various cleavage methods.

#### 6.10.3. Purification of the Peptides

Purification of the synthesized peptides can be carried out by standard methods including chromatography (e.g., ion exchange, affinity, and sizing column chromatography, high performance liquid chromatography

(HPLC)), centrifugation, differential solubility, or by any other standard technique.

6.10.4. Conjugation of Peptides to Other Molecules

5

The peptides of the present invention may be linked to other molecules (e.g., a detectable label, a molecule facilitating adsorption to a solid substratum, or a toxin, according to various embodiments of the invention) by methods 10 that are well known in the art. Such methods include the use of homobifunctional and heterobifunctional cross-linking molecules.

The homobifunctional molecules have at least two reactive functional groups, which are the same. The reactive 15 functional groups on a homobifunctional molecule include, for example, aldehyde groups and active ester groups.

Homobifunctional molecules having aldehyde groups include, for example, glutaraldehyde and subaraldehyde. The use of glutaraldehyde as a cross-linking agent was disclosed by 20 Poznansky et al., 1984, Science 223:1304-1306.

Homobifunctional molecules having at least two active ester units include esters of dicarboxylic acids and N-hydroxysuccinimide. Some examples of such N-succinimidyl esters include disuccinimidyl suberate and dithio-bis- 25 (succinimidyl propionate), and their soluble bis-sulfonic acid and bis-sulfonate salts such as their sodium and potassium salts. These homobifunctional reagents are available from Pierce, Rockford, Illinois.

The heterobifunctional molecules have at least two 30 different reactive groups. Some examples of heterobifunctional reagents containing reactive disulfide bonds include N-succinimidyl 3-(2-pyridyl-dithio)propionate (Carlsson et al., 1978, Biochem J. 173:723-737), sodium S-4-succinimidylloxycarbonyl-alpha-methylbenzylthiosulfate, and 35 4-succinimidylloxycarbonyl-alpha-methyl-(2-pyridyldithio)toluene. N-succinimidyl 3-(2-pyridyldithio)propionate is preferred. Some examples of

heterobifunctional reagents comprising reactive groups having a double bond that reacts with a thiol group include succinimidyl 4-(N-maleimidomethyl)cyclohexahe-1-carboxylate and succinimidyl m-maleimidobenzoate.

5 Other heterobifunctional molecules include succinimidyl 3-(maleimido)propionate, sulfosuccinimidyl 4-(p-maleimido-phenyl)butyrate, sulfosuccinimidyl 4-(N-maleimidomethyl-cyclohexane)-1-carboxylate, maleimidobenzoyl-N-hydroxy-succinimide ester. The sodium sulfonate salt of 10 succinimidyl m-maleimidobenzoate is preferred. Many of the above-mentioned heterobifunctional reagents and their sulfonate salts are available from Pierce.

Additional information regarding how to make and use these as well as other polyfunctional reagents may be 15 obtained from the following publications or others available in the art: Carlsson et al., 1978, Biochem. J. 173:723-737; Cumber et al., 1985, Methods in Enzymology 112:207-224; Jue et al., 1978, Biochem 17:5399-5405; Sun et al., 1974, Biochem. 13:2334-2340; Blattler et al., 1985, Biochem. 20 24:1517-152; Liu et al., 1979, Biochem. 18:690-697; Youle and Neville, 1980, Proc. Natl. Acad. Sci. USA 77:5483-5486; Lerner et al., 1981, Proc. Natl. Acad. Sci. USA 78:3403-3407; Jung and Moroi, 1983, Biochem. Biophys. Acta 761:162; Caulfield et al., 1984, Biochem. 81:7772-7776; Staros, 1982, 25 Biochem. 21:3950-3955; Yoshitake et al., 1979, Eur. J. Biochem. 101:395-399; Yoshitake et al., 1982, J. Biochem. 92:1413-1424; Pilch and Czech, 1979, J. Biol. Chem. 254:3375-3381; Novick et al., 1987, J. Biol. Chem. 262:8483-8487; Lomant and Fairbanks, 1976, J. Mol. Biol. 104:243-261; Hamada 30 and Tsuruo, 1987, Anal. Biochem. 160:483-488; Hashida et al., 1984, J. Applied Biochem. 6:56-63.

Additionally, methods of cross-linking are reviewed by Means and Feeney, 1990, Bioconjugate Chem. 1:2-12.

35 **6.10.4.1. Biotinylation of Peptides**

Methods of biotinyling peptides are well known in the art. Any convenient method may be employed in the

practice of the invention. For example, the following procedure was used. Ten micrograms of peptide was dissolved in 100  $\mu$ l of 0.1 % acetic acid. PBS (900 $\mu$ l) and 3.3 mg of biotin-LC-NHS (Pierce, Rockford, IL) was added. Following 5 incubation for 30 minutes at room temperature the biotinylated peptides were purified over a Superose 12 column (Pharmacia, Piscataway, NJ).

#### 6.10.5. Synthetic Peptides

10 Tables 19, 20 and 21 provide the primary structure for various synthetic peptides manufactured in the practice of the present invention.

15

Table 19

Seq ID No	Peptide name	Sequence
20	ELAN005	H <sub>2</sub> N-C-K(dns) - FITKALGISYGRKKRRQRRPPQGSQTHQVSLSKQ-CONH <sub>2</sub>
167	ELAN006 FITC- ELAN006	Ac-CLNGGVKMYVESVDRYVC-CONH <sub>2</sub> Ac-CLNGGVK (FITC) MYVESVDRYVC-CONH <sub>2</sub>
193	ELAN006ii ELAN007	H <sub>2</sub> N-C-K(dns) -RLNGGVSMYVESVDRYVCR-CONH <sub>2</sub> H <sub>2</sub> N-RIAGLPWYRCRTVAFETGMQNTQLCSTIVQLSFTPSEE-COOH
25	ELAN007ii bZElan008 (P31)	H <sub>2</sub> N-KKRIAGLPWYRCRTVAFETGMQNTQLCSTIVQLSFTPSEE-CONH <sub>2</sub> biotin-K(dns) SARDSGPAEDGSRAVRLNGVENANTRKSSR SNPRGRRHP-COOH
168	bZElan009 ELAN010	biotin-K(dns) SSADAECAGSLLWWGRQNNNSCGSPTKKH LKHRNRSQTSSSHG-COOH H <sub>2</sub> N-REFAERRLWGDDLSWRDAEGCGPTPSNRAVKHRKPRPR SPAL-COOH
30	bZElan010 ELAN012	biotin-K(dns) REFAERRLWGDDLSWRDAEGCGPTPSNR AVKHRKPRPRSPAL-COOH H <sub>2</sub> N- SGSHSGGMNRAYGDVFRELDRWYATSHTRPTPQLPRGPN-COOH
35	bELAN012 ZElan012	biotin- SGSHSGGMNRAYGDVFRELDRWYATSHTRPTPQLPRGPN-COOH H <sub>2</sub> N- K(dns) SGSHSGGMNRAYGDVFRELDRWYATSHTRPTPQLP RGPN-COOH

	249	ELAN013	H <sub>2</sub> N- SGSPCCGGSWGRFMQGGLFGGRTDGCAGHRNRTSASLEPPSSD
	250	ELAN014	H <sub>2</sub> N- Y-CONH <sub>2</sub> SHSGGMNRAYGDVFRELDRWNATSHHTRPTPQLPRGPNS-
5		bZElan014	CONH <sub>2</sub> biotin- K(dns) SHSGGMNRAYGDVFRELDRWNATSHHTRPTPQLPRG
		ZEElan014	PNS-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) SHSGGMNRAYGDVFRELDRWNATSHHTRPTPQLPRG
		ZEElan015 (DCX11)	PNS-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) SQGSKQCMQYRTGRLTVGSEYGCGMNPARHATPAYPA
10		ZEElan016 (SNi10)	RLLPRYR-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) RVGQCTDSDVRRPWARSCAHQGCGAGTRNSHGCITRP
		bZEElan017	LRQASAH-CONH <sub>2</sub> biotin-K(dns) SGSGRVGQCTDSDVRRPWARSCA-CONH <sub>2</sub>
		ZEElan017	H <sub>2</sub> N-K(dns) RVGQCTDSDVRRPWARSCA-CONH <sub>2</sub>
		ZEElan018 (PAX2)	H <sub>2</sub> N- K(dns) STPPSREAYSRPYSVDSDSDTNAKHSSHNRRLRTRSR
15		ZEElan019 (5PAX5)	PNG-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) RGSTGTAGGERSGVNLHTRDNASGSGFKPWYPSNRG
		ZEElan020 (CY09)	HK-CONH <sub>2</sub> H <sub>2</sub> N-K(dns) SGSGLYANPGMYSRLHSPA-CONH <sub>2</sub>
	20	bZEElan020 (CY09)	biotin-K(dns) SGSGLYANPGMYSRLHSPA-CONH <sub>2</sub>
		ZEElan021 (HAX42)	H <sub>2</sub> N- K(dns) SDHALGTNLRSDNAKEPGDYNCCGNGNSTGRKVFNRR
		ZEElan022 (SNi34)	RPSAIPT-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) SPCGGSWGRFMQGGLFGGRTDGCAGHRNRTSASLEPP
25		ZEElan023 (DCX8)	SSDY-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) RYKHDIGCDAGVDKSSSVRGCGAHSSPPRAGRGPR
		ZEElan024 (P31)	GTMVSRL-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRR
		ZEElan025 (DAB10)	HPGG-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) SKSGEGGDSSRGETGWARVRSHAMTAGRFRWYNQLPS
30		ZEElan026 (PAX2/control)	DR-CONH <sub>2</sub> H <sub>2</sub> N- K(dns) SEANLDGRKSRYSSPRRNSSTRPRTSPNSVHARYPST
		bELAN027 (PAX2)	DHD-CONH <sub>2</sub> biotin- SGSGSTPPSREAYSRPYSVDSDSDTNAKHSSHNRRLRTRSRPN
35	251	18C21	G-CONH <sub>2</sub> H <sub>2</sub> N-DTNAKHSSHNRRLRTRSRPN-CONH <sub>2</sub>
		Fmoc-	Fmoc-K(dns) RVGQCTDSDVRRPWARSCAHQG-COOH
		Z16N23	
	252	16C23	H <sub>2</sub> N-CGAGTRNSHGCITRPLRQASAHG-CONH <sub>2</sub>

	Z16C23	H <sub>2</sub> N-K (dns) CGAGTRNSHGCITRPLRQASAHG-CONH <sub>2</sub>
	ZEelan028 (P31 fragment)	H <sub>2</sub> N-K (dns) ENANTRKSSRSNPRGRRHPG-CONH <sub>2</sub>
5	ZEelan029 (P31 fragment)	H <sub>2</sub> N-K (dns) TRKSSRSNPRG-CONH <sub>2</sub>
	ZEelan030 (P31 fragment)	H <sub>2</sub> N-K (dns) ENANTRKSSRSNPRG-CONH <sub>2</sub>
	ZEelan031 (P31 fragment)	H <sub>2</sub> N-K (dns) TRKSSRSNPRGRRHPG-CONH <sub>2</sub>
10	ZEelan032 (PAX2 fragment)	H <sub>2</sub> N-K (dns) TNAKHSSHNRRLRTRSRPN-CONH <sub>2</sub>
	ZEelan033 (PAX2 fragment)	H <sub>2</sub> N-K (dns) TNAKHSSHNRRLRTR-CONH <sub>2</sub>
	ZEelan034 (PAX2 fragment)	H <sub>2</sub> N-K (dns) SSHNRLRTRSRPN-CONH <sub>2</sub>
15	ZEelan035 (PAX2 fragment)	H <sub>2</sub> N-K (dns) SSHNRLRTR-CONH <sub>2</sub>
	ZEelan036 (SNi10 fragment)	H <sub>2</sub> N-K (dns) VRRPWARSCAHQGCGAGTRNS-CONH <sub>2</sub>
20	ZEelan037 (SNi10 fragment)	H <sub>2</sub> N-K (dns) CTDSDVRRPWARSC-CONH <sub>2</sub>
	ZEelan038 (PAX2/con trol)	H <sub>2</sub> N- K (dns) SRANTDGRKSRYSSPRRNSSTEPRLSPNSVHARYPST DHD-CONH <sub>2</sub>
	ZEelan039 (P31 fragment)	H <sub>2</sub> N-K (dns) ENANTRKSSR-CONH <sub>2</sub>
25	ZEelan040 (P31 fragment)	H <sub>2</sub> N-K (dns) SNPRGRRHPG-CONH <sub>2</sub>
	ZEelan041 (P31 fragment)	H <sub>2</sub> N-K (dns) ENANT-CONH <sub>2</sub>
	ZEelan042 (P31 fragment)	H <sub>2</sub> N-K (dns) ANTRKS-CONH <sub>2</sub>
30	ZEelan043 (P31 fragment)	H <sub>2</sub> N-K (dns) TRKSS-CONH <sub>2</sub>
	ZEelan044 (P31 fragment)	H <sub>2</sub> N-K (dns) RKSSR-CONH <sub>2</sub>
35	ZEelan045 (P31 fragment)	H <sub>2</sub> N-K (dns) KSSRSN-CONH <sub>2</sub>

	ZElan046 (P31 fragment) ZElan047 (P31 fragment) ZElan048 (P31 fragment) ZElan049 (P31 fragment) ZElan050 (P31 fragment) ZElan051 (HepC) ZElan052 (HepC) ZElan053 (P31 analog) ZElan054 (P31 analog) ZElan055 (PAX2 fragment) ZElan056 (PAX2 fragment) ZElan057 (PAX2 fragment) ZElan058 (PAX2 fragment) ZElan059 (PAX2 analog) ZElan060 (HAX42 fragment) ZElan061 (HAX42 fragment) ZElan062 (HAX42 fragment) ZElan063 (HAX42 fragment) ZElan064 (HAX42 fragment)	H <sub>2</sub> N-K(dns)SSRSNPG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)RSNPRG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SNPRG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)PRGRRH-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)RRHPG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)KSSRGN-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)KTSERSQPRGRRQPG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)TrKSSrSNPrGrrHPG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)TRKSSrSNPRGrRHPG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)TNAKHSSHN-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)RRLRTRSRPN-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)RRLRTRSR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)RRLRTR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)rrLrTrSrPN-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAKEPGDYNCCGNG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)GDYNCCGNGNSTGRKVFNRRRPSAIFT-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAKEPG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)GDYNCCGNGNSTG-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)RKVFNRRRPSAIFT-CONH <sub>2</sub>
5		
10		
15		
20		
25		
30		
35		

	ZElan065 (HAX42 fragment) ZEElan066 (HAX42 fragment) ZEElan067 (HAX42 fragment) Elan018 (PAX2 no dns) Elan021 (HAX42 no dns) ZEElan070 (HAX42 fragment) ZEElan071 (HAX42 fragment) ZEElan072 (HAX42 fragment) ZEElan073 (PAX2 fragment) ZEElan074 (PAX2 fragment) ZEElan075 (PAX2 fragment) ZEElan076 (PAX2 fragment) ZEElan077 (PAX2 fragment) ZEElan078 (PAX2 fragment) ZEElan079 (PAX2 fragment) ZEElan080 (PAX2 fragment) ZEElan081 (PAX2 fragment) ZEElan082 (PAX2 fragment) Elan035	H <sub>2</sub> N-K(dns)RKVFNRRRPS-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)NRRRPSAIPT-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)NRRRPS-CONH <sub>2</sub> H <sub>2</sub> N- STPPSREAYSRPYSVDSDDSDTNAKHSSHNRRLRTRSRPNG- CONH <sub>2</sub> H <sub>2</sub> N-SDHALGTNLRSNAKEPGDYNCCGNGNSTGRKVFNRRRPS AIPT-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SDHALGTNLRSNAKEPGDYNCCGNGNST- CONH <sub>2</sub> H <sub>2</sub> N-K(dns)NLRSDNAKEPGDYNCCGNGNSTGRKVFNR- CONH <sub>2</sub> H <sub>2</sub> N-K(dns)PGDYNCCGNGNSTGRKVFNRRPSAIPT-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)ASHNRRLRTR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SAHNRRRLRTR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SSANRRRLRTR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SSHARRRLRTR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SSHNARLRTR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SSHNRALRTR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SSHNRARTR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SSHNRRLATR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SSHNRRLRAR-CONH <sub>2</sub> H <sub>2</sub> N-K(dns)SSHNRRLRTA-CONH <sub>2</sub> H <sub>2</sub> N-SSHNRRLRTR-CONH <sub>2</sub>
--	---	---

5	ZElan083 (PAX2/control)	H <sub>2</sub> N-K(dns) GRNHDVVSSNTHKSYRSPRSASYPRLSNDRTDRTEPA
	ZElan084 (PAX2/control)	PSS-CONH <sub>2</sub> H <sub>2</sub> N-K(dns) RNTRNKTSLSANPHRSHR-CONH <sub>2</sub>
	Elan032Z (PAX2 fragment)	H <sub>2</sub> N-TNAKHSSHNRRLRTRSRPN K(dns)-CONH <sub>2</sub>
	Elan057Z (PAX2 fragment)	H <sub>2</sub> N-RRLRTRSRK(dns)-CONH <sub>2</sub>

10

TABLE 20		
	Name	Description
15	ZElan087	HAX42-1 (20 mer)
	ZElan088	HAX42-2 (20 mer)
	ZElan089	HAX42-3 (15 mer)
	ZElan090	HAX42-4 (15 mer)
	ZElan091	HAX42-5 (14 mer)
	ZElan092	HAX42-6 (10 mer)
	ZElan093	HAX42-7 (10 mer)
20	ZElan100	P31 16 mer cyclic
	ZElan101	P31 16 mer cyclic D form
25	ZElan103	PAX2 15 mer cyclic
	ZElan103A	PAX2 15 mer cyclic (internal)
30	ZElan104	PAX2 15 mer cyclic (internal)
	ZElan105	PAX2 Ala Scan 1
35	ZElan106	PAX2 Ala Scan 2
	ZElan107	PAX2 Ala Scan 3
	ZElan108	PAX2 Ala Scan 4
	ZElan109	PAX2 Ala Scan 5
	ZElan110	PAX2 Ala Scan 6
	ZElan111	PAX2 Ala Scan 7
	ZElan112	PAX2 Ala Scan 8

ZElan113	PAX2 Ala Scan 9	H <sub>2</sub> N-K(dns) TNAKHSSHARRLRTR	
ZElan114	PAX2 Ala Scan 10	H <sub>2</sub> N-K(dns) TNAKHSSHNRALRTR	
ZElan115	PAX2 Ala Scan 11	H <sub>2</sub> N-K(dns) TNAKHSSHNRALRTR	
ZElan116	PAX2 Ala Scan 12	H <sub>2</sub> N-K(dns) TNAKHSSHNRARTR	
ZElan117	PAX2 Ala Scan 13	H <sub>2</sub> N-K(dns) TNAKHSSHNRRLATR	
5	ZElan118	PAX2 Ala Scan 14	H <sub>2</sub> N-K(dns) TNAKHSSHNRRLRAR
ZElan119	PAX2 Ala Scan 15	H <sub>2</sub> N-K(dns) TNAKHSSHNRRLRTA	
ZElan123	PAX2 15 mer cyclic D form	H <sub>2</sub> N-K(dns) Lys-TNAKHSSHNrLrTr	
10	ZElan124	PAX2 15 mer D form	H <sub>2</sub> N-K(dns) TNAKHSSHNrLrTr
ZElan125	PAX2 10 mer cyclic	H <sub>2</sub> N-K(dns) Lys-SSHNRRLRTR	
ZElan126	PAX2 10 mer cyclic D form	H <sub>2</sub> N-K(dns) Lys-SSHNrLrTr	
ZElan127	PAX2 10 mer cyclic	H <sub>2</sub> N-K(dns) Lys-TNAKHSSHNR	
15	ZElan128	PAX2 10 mer cyclic D form	H <sub>2</sub> N-K(dns) Lys-TNAKHSSHNr
ZElan129	PAX2 15 mer	H <sub>2</sub> N-K(dns) TNAKHSSHNRRLRTR	
ZElan130	HAX42 14 mer Ala Scan 1	H <sub>2</sub> N-K(dns) AGDYNCCGNGNSTG	
20	ZElan131	HAX42 14 mer Ala Scan 2	H <sub>2</sub> N-K(dns) PADYNCCGNGNSTG
ZElan132	HAX42 14 mer Ala Scan 3	H <sub>2</sub> N-K(dns) PGAYNCCGNGNSTG	
ZElan133	HAX42 14 mer Ala Scan 4	H <sub>2</sub> N-K(dns) PGDANCCGNGNSTG	
ZElan134	HAX42 14 mer Ala Scan 5	H <sub>2</sub> N-K(dns) PGDYACCGNGNSTG	
25	ZElan135	HAX42 14 mer Ala Scan 6	H <sub>2</sub> N-K(dns) PGDYNACGNGNSTG
ZElan136	HAX42 14 mer Ala Scan 7	H <sub>2</sub> N-K(dns) PGDYNCAAGNSTG	
ZElan137	HAX42 14 mer Ala Scan 8	H <sub>2</sub> N-K(dns) PGDYNCCANGNSTG	
30	ZElan138	HAX42 14 mer Ala Scan 9	H <sub>2</sub> N-K(dns) PGDYNCCGAGNSTG
ZElan139	HAX42 14 mer Ala Scan 10	H <sub>2</sub> N-K(dns) PGDYNCCGNANSTG	
ZElan140	HAX42 14 mer Ala Scan 11	H <sub>2</sub> N-K(dns) PGDYNCCGNGASTG	
ZElan141	HAX42 14 mer Ala Scan 12	H <sub>2</sub> N-K(dns) PGDYNCCGNGNATG	
35	ZElan142	HAX42 14 mer Ala Scan 13	H <sub>2</sub> N-K(dns) PGDYNCCGNGNSAG
ZElan143	HAX42 14 mer Ala Scan 14	H <sub>2</sub> N-K(dns) PGDYNCCGNGNSTA	

GST fusion proteins of GIT peptides are shown in  
Table 21.

5

10

15

20

25

30

35

Table 21

Source	Clone #	GST Fusion Sequence	SEQ ID NO.
DCX11	98	gst-SQGSKQCMQYRTGRLTVGSEYCGGMNPARTHATPAYPARLLPRYR	213
HAX42	99	gst-SDHALGTNLRSNDNAKEPGDYNCCGNGNSTGRKVFNRRPSAIFT	214
SNi34	100	gst-SPCGGGSWGRFMQGGLFGGRTDGCGAHRNRTSASLEPPSSDY	215
SPAX5	97	gst-RGSTGTAGGERSGVNLHTRDNASGSGFKPWYPSNRGHK	216
SNi28	84	gst-SHSGGMNRAYGDVFRELDRWNATSHHTRPTQLPRGPN	217
SNi28	85	gst-SHSGGMNRAY	218
SNi28	86	gst-GDVFRELDR	219
SNi28	87	gst-WNATSHHTRP	220
SNi28	88	gst-TPQLPRGPN	221
SNi28	89	gst-GDVFRELDRWNATSHHTRP	222
SNi28	90	gst-WNATSHHTRPTQLPRGPN	223
SNi28	91	gst-GDVFRELDRWNATSHHTRPTQLPRGPN	224
SNi28	92	gst-SHSGGMNRAYGDVFRELDRWNATSAATRPTQLPRGPN	225
P31	93	gst-SARDSGPAEDGSRAVRLNGVENANTRKSSRSNIPRGRRHP	226
P31	101	gst-SARDSGPAEDGSRAVRLNG	227
P31	102	gst-DGSRAVRLNGVENANTRKSSR	228
P31	103	gst-ENANTRKSSRSNPRGRRHP	229
P31	110	gst-ENANTRKSSR	230

P31	111	gst-RKSSRSNPRG	
P31	112	gst-SNPRGRRHP	232
P31	119	gst-TRKSSRSNPRG	233
PAX2	94	gst-STPPSREAYSRPYSVSDSDTNAKHSSHNRRLTRSRPN	234
PAX2	104	gst-STPPSREAYSRPYSVSDSDSD	235
PAX2	105	gst-SRPYSVSDSDTNAKHSSHNR	236
PAX2	106	gst-TNAKHSSHNRRLTRSRPN	237
PAX2	113	gst-TNAKHSSHIN	238
PAX2	114	gst-SSHNRRLRTR	239
PAX2	115	gst-RRLTRSRPN	240
SNi10	96	gst-RVGQCTDSVRRPWAHQQCGAGTRNSHGCTITRPLRQASAH	241
SNi10	116	gst-RVGQCTDSVRRPWAHSCA	242
SNi10	117	gst-VRRPWAHSCAHQGGAGTRNS	243
SNi10	118	gst-GTRNSHGCITRPLRQASAH	244
DCX8	95	gst-RYKHDIGCDAGVDKSSSVRGCCGAHSSPPRAGRGPRGTMVSRL	245
DCX8	107	gst-RYKHDIGCDAGVDKSSSVRGCCG	246
DCX8	108	gst-GCDAGVDKSSSVRGCCGAHSSPPRA	247
DCX8	109	gst-GAHSPPRAGRGPRGTMVSRL	248

**6.10.6. Peptide Stability**

The relative stability for ZElan031, ZElan053 and ZElan054 was determined in simulated intestinal fluid (SIF) SIF was made by dissolving 100mg of pancreatin (Sigma cat#P-5 1625, lot# 122H0812) in 8.4ml of phosphate stock solution, adjusting the pH to 7.5 with 0.2N NaOH and adjusting the volume to 10ml with water.

Peptide (3.25mg) was dissolved in 3.25 ml of 10,000 fold diluted SIF solution at 37°C. Aliquots (0.7ml) of the 10 digestion solution were then withdrawn at <1min, 1h, 3h, and 21h or 24h. The samples were quickly passed through a syringe filter (Millipore Millex-GV 0.22μm, part# SLGV025LS, lot# H2BM95250) and 300μL of the filtered solution was immediately injected onto a Hewlett-Packard HPLC system equipped with a 15 C-8 column (Applied Biosystems column and guard column: column- p/n 0711-0023 Spheri-5 ODS 5μm, 220x4.6mm). The products were eluted at 1.5ml/min using an acetonitrile-water gradient. The major fluorescent peaks were collected, lyophilized and identified by MS analysis.

20 The HPLC gradient used was:

Time (min)	Solvent Mixture	
0	95% H <sub>2</sub> O-5% acetonitrile (0.1%TFA)	
5	95% H <sub>2</sub> O-5%acetonitrile (0.1%TFA)	
35	85% H <sub>2</sub> O-15% acetonitrile (0.1%TFA)	linear solvent change
25 40	0% H <sub>2</sub> O-100% acetonitrile (0.1%TFA)	"
45	95% H <sub>2</sub> O-5% acetonitrile (0.1%TFA)	"
52	95% H <sub>2</sub> O-5%acetonitrile (0.1%TFA)	"

As shown in Table 22, the relative stability (to SIF) for the three peptides was found to be 30 ZElan053>ZElan054>ZElan031. Enzymatic cleavage of the peptide was found to occur at arginine and/or lysine as expected. The replacement of L-amino acids with their D-amino acid analogs significantly reduced the rate of proteolysis at these residues.

35

TABLE 22

<u>Peptide</u>	<u>Percent Remaining at:</u>				<u>Rel. Stab.</u>
	<u>1 m</u>	<u>1 h</u>	<u>3 h</u>	<u>24 h</u>	
5 ZElan031	100	38.7	0	0	3
ZElan054	97.4	58.2	11.6	2.7	2
ZElan053	100	98.3	98.1	94.0	1

10 7. CHARACTERIZATION OF PEPTIDE-COATED PARTICLESBinding of Peptide-Coated PLGA Nanoparticles  
to Fixed Caco-2 Cells

15 Binding of nanoparticles coated with targeting peptides to fixed Caco-2 cells was investigated using an ELISA assay based on reaction of antibody with the dansyl moiety present on the peptides. Isoelectric points of selected synthetic peptides are shown in Table 23 (corresponding SEQ ID NOS. are shown in Table 7). Corresponding dansylated synthetic GIT binding peptides are given in Table 24.

TABLE 23

<u>Peptide</u>	<u>Sequence</u>	<u>pI</u>
P31	SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRRHP	12.26
25 5PAX5	RGSTGTAGGERSGVNLNLHTRDNASGSGFKPWYPSNRGHK	11.49
SNi10	RVGQCTSDVRRPWARSCAHQGCGAGTRNSHGCITRPLRQASAH	10.45
SNi34	SPCGGSWGRFMQGGLFGGRTDGCAGHRNRTSASLEPPSSDY	8.25
DCX11	SQGSKQCMQYRTGRLTVGSEYGCGMNPARHATPAYPARLLPRYR	10.44
DCX8	RYKHDIGCDAGVDKSSSVRGCGAHSSPPRAGRGPRGTMVSRL	11.03
HAX42	SDHALGTNLRSDNAKEPGDYNCCGNGNSTGRKVFNRRRPSAIP	9.62
30 PAX2	STPPSREAYSRPYSVDSDSDTNAKHSSHNRRLRTRSRPN	11.26

TABLE 24

<u>Peptide</u>	<u>Sequence</u>
P31	H <sub>2</sub> N-K(dns)SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRRHPGG-CONH <sub>2</sub>
SPAX5	H <sub>2</sub> N-K(dns)RGSTGTAGGERSGVNLHTRDNASGSGFKPWYPSNRGHK-CONH <sub>2</sub>
5 SNi10	H <sub>2</sub> N-K(dns)RVGQCTDSVRRPWARSCAHQGCGAGTRNSHGCITRPLRQASAH-CONH <sub>2</sub>
SNi34	H <sub>2</sub> N-K(dns)SPCGGSWGRFMQGGLFGGRTDGCGAHRNRTSASLEPPSSDY-CONH <sub>2</sub>
DCX11	H <sub>2</sub> N-K(dns)SQGSKQCMQYRTGRLTVGSEYGGCGMNPARHATPAYPARLLPRYR-CONH <sub>2</sub>
DCX8	H <sub>2</sub> N-K(dns)RYKHDIGCDAGVDKSSSVRGCGAHSSPPRAGRGPRGTMVSRL-CONH <sub>2</sub>
HAX42	H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAKEPGDYNCCGNGNSTGRKVFNRRRPSAIP-CONH <sub>2</sub>
PAX2	H <sub>2</sub> N-K(dns)STPPSREAYSRPYSVDSDSDTNAKHSSHNRRLRTRSRPN-CONH <sub>2</sub>
10 DAB10	H <sub>2</sub> N-K(dns)SKSGEGGDSSRGETGWARVRSHAMTAGRFRWYNQLPSDR-CONH <sub>2</sub>

Method:

Confluent Caco-2 monolayers grown in 96-well plates (p38) were fixed and treated with 0.1% phenylhydrazine before blocking with 0.1% BSA in PBS. Control and dansyl peptide-coated nanoparticles were resuspended in sterile water at 10mg/ml and stirred with a magnet for 1h at room temperature. Samples consisted of: (1) blank nanoparticle control, (2) scrambled PAX2-coated nanoparticles, (3) PAX2-coated nanoparticles, (4) HAX42-coated nanoparticles, (5) PAX2/HAX42-coated nanoparticles, and (6) 8 peptide-coated nanoparticles.

Nanoparticles were added to the cells at 10mg/ml in 100 $\mu$ l 1%BSA-PBS (no Tween80 is used in this assay) and 2-fold serially-diluted. The 96-well plates were incubated for 1h at room temperature. The plates were washed 5 times with 1%BSA-PBS and 100 $\mu$ l of anti-dansyl antibody (Cytogen DB3-226.3; 0.5  $\mu$ g/ml; batch May 1997) was added per well and the plates incubated 1h at room temperature. The wells were washed 5 times with 1%BSA-PBS; 100 $\mu$ l of goat anti-mouse  $\lambda$ :HRP antibody (Southern Biotechnology CN. 1060-05; 1:10,000) was added per well, and the plates incubated 1h at room temperature. After washing 5 times with 1%BSA-PBS, 100 $\mu$ l of TMB peroxidase substrate (KPL CN. 50-76-00) was added to the wells and the optical density at 650nm was measured after 15 minutes.

As shown in Figures 13A-B, a decreasing anti-dansyl ELISA response was observed for nanoparticles coated with PAX2, HAX2, PAX2+HAX2, and a mixture of 8 targeting peptides, when decreasing amounts of the nanoparticles were applied to 5 fixed Caco-2 cells. No concentration effect was observed for blank nanoparticles or nanoparticles coated with a scrambled version of PAX2 peptide. Nanoparticles coated with PAX2, HAX2, PAX2+HAX2, and the 8 peptide mix, showed increased response relative to blank nanoparticles or nanoparticles 10 coated with a scrambled version of PAX2 peptide. The OD values were low relative to those normally observed for GST-peptide fusion binding to fixed Caco-2 cells.

Table 25 below shows the insulin potency and level 15 of peptides coated onto the particles (measured by fluorescense) for formulation 1 particles (formulation by the coacervation method given below).

20

Table 25

	Peptide	Blend	
		Insulin mg/g	Peptide μl/mg
25	PAX2	60.7	3.51
	HAX42	55.9	2.93
	PAX2 SCRAMBLED	57.7	1.26
	P31	67.0	1.22
	5PAX5	52.7	2.83
	SNI10	59.5	1.75
	SNI34	61.5	4.03
	DCX8	59.1	1.87
	DAB10	55.9	1.99

30

#### ELISA of dansylated peptides and insulin coated PLGA particles

The standard ELISA procedure was modified as 35 follows. Peptides and particles were diluted to an appropriate concentration in PBS containing 1%BSA (particles were sonicated to achieve a homogeneous solution), titered

and incubated one hour at room temperature. Following five washes with PBS containing 1%BSA, an in-house IgG1 $\lambda$  anti-dansyl monoclonal antibody was added (diluted to 1 $\mu$ g/ml in 1%BSA-PBS) and the plates were incubated for one hour. After 5 five more washes goat anti-mouse  $\lambda$ -HRP was added (Southern Biotechnology Associates Inc., Birmingham, AL, diluted 1:10,000 in 1%BSA-PBS) and the plates were incubated one hour. After five washes, plates were developed with TMB peroxidase substrate (Kirkegaard and Perry, Gaithersburg, MD). 10 All data is presented with background binding subtracted. Tween 20 was not added to the diluent or the washes when insulin coated PLGA particles were included in the assay.

Figures 14A-14B show the binding of the dansylated 15 peptide SNi10 to hSI and BSA.

#### 8. BINDING OF SYNTHETIC PEPTIDES AND PEPTIDE-COATED PARTICLES TO S100 AND P100 FRACTIONS DERIVED FROM CACO-2 CELLS

20

8.1. Detection of Binding to Membrane (P100) and Cytosolic (S100) fractions

Caco-2 cell membrane (P100) and cytosolic (S100) fractions were prepared using a modification of the method 25 described in Kinsella, B. T., O'Mahony, D. J. and G. A. FitzGerald, 1994, J. Biol. Chem. 269(47): 29914-29919. Confluent Caco-2 cell monolayers (grown in 75 cm<sup>2</sup> flasks for up to 1 week at 37°C and 5% CO<sub>2</sub>) were washed twice in Dulbecco's PBS (DPBS) and the cells were harvested by 30 centrifugation at 1000 rpm after treatment with 10 mM EDTA-DPBS. The cells were washed 3 times in DPBS and the final cell pellet was resuspended in 3 volumes of ice cold HED buffer (20 mM HEPES (pH 7.67), 1 mM EGTA, 0.5 mM dithiothreitol, 1 mM phenylmethylsulphonyl fluoride (PMSF)). 35 The cells were allowed to swell for 5 min on ice prior to homogenization for 30 sec. The homogenates were centrifuged at 40,000 rpm for 45 min at 4°C. The supernatant (S100) was

removed and the pellet (P100) was resuspended in HEDG buffer (20 mM HEPES (pH 7.67), 1 mM EGTA, 0.5 mM dithiothreitol, 100 mM NaCl, 10% glycerol, 1 mM PMSF). Protein concentrations were determined using the Bradford assay (Bradford, M. M., 5 1976, *Anal. Biochem.* 72: 248-254).

Binding of peptide and/or peptide-coated PLGA particles to membrane (P100) and cytosolic (S100) fractions was assessed by detection of the dansyl moiety incorporated in the peptide. Costar ninety six well ELISA plates were 10 coated with S100 and P100 fractions (100 µg/ml in 0.05 M NaHCO<sub>3</sub>) overnight at 4°C. The plates were blocked with 0.5% bovine serum albumin in DPBS for 1 h at room temperature and washed 3 times in 1% BSA-DPBS. Peptide-coated particles or peptides were dispersed in the same buffer and added to the 15 plates at concentrations in the range 0.0325 - 0.5 mg/well. After 1 h at room temperature the plates were washed 5 times in 1% BSA-DPBS and 100 µl of anti-dansyl antibody (Cytogen DB3-226.3; 0.5 µg/ml) was added per well. The plates were 20 incubated for 1 h at room temperature. The wells were washed 3 times in 1% BSA-DPBS and 100 µl of goat anti-mouse IgGλ:HRP antibody (Southern Biotechnology 1060-05; 1:10,000) was added per well. The plates were incubated for 1 h at room temperature. After washing 3 times in 1% BSA-DPBS 100 µl of 25 TMB substrate (3,3',5',5-tetramethylbenzidine; Microwell Peroxidase Substrate System (Kirkegaard and Perry Laboratories 50-76-00)) was added and the optical density was measured at 650 nm at various time intervals.

#### 8.2. Binding of Peptide-Coated PLGA particles

30 A novel assay system is provided by the instant invention for detection of binding of peptide-coated PLGA particles to membrane (P100) and cytosolic (S100) fractions derived from live Caco-2 cells. The absorbance readings obtained using this assay system were substantially higher 35 than those obtained using similar peptide-coated PLGA particle concentrations on fixed Caco-2 cells. This greater sensitivity together with the derivation of the S100 and P100

fractions from live Caco-2 cells suggests that this assay may be the assay system of choice for detection of peptide-coated PLGA particle binding. The assay was concentration dependent and peptide/particle correlation permitted differentiation 5 between specific and non-specific binding interactions.

Binding of peptide-coated PLGA particles was assessed using S100 and P100 fractions derived from live Caco-2 cells as described above. The fractions were coated onto 96-well plates at 10 $\mu$ g/well in 0.05 M NaHCO<sub>3</sub>, and peptide-coated PLGA 10 particles were assayed by ELISA at concentrations in the range 0.0325 - 0.5 mg/well.

Figures 15A and 15B illustrate the data obtained on S100 and P100 fractions respectively for particles coated with no peptide, scrambled PAX2 (control), P31 D-Arg 16-mer 15 (ZElan053), HAX42, PAX2 and HAX42/PAX2. Using particle concentrations of 0.0325 - 0.5 mg/well all test peptide-coated PLGA particles exhibited greater binding to both the S100 and P100 fractions than the scrambled PAX2 coated control particles. All particles except P31 D-Arg 16-mer 20 (ZElan053) exhibited greater binding to the P100 fraction than the S100 fraction. Greater binding of the P31 D-Arg 16-mer (ZElan053) coated particles to the S100 fraction may be indicative of non-specific binding due to the D-Arg modification of the P31 peptide (SEQ ID NO:43).

25 Binding of PLGA particles coated with varying concentrations of PAX2 peptide ranging from 0.05 - 5.0 mg/g was assessed using a) fixed Caco-2 cells (P35) and b) S100 and P100 fractions (Caco-2 P33). The particles were assayed at concentrations in the range 0.03125 - 0.0625 mg/well.

30 Using a particle concentration of 0.0625 mg/well, all PAX2 coated particles except those coated at 0.05 mg/g exhibited greater binding to fixed Caco-2 cells than the scrambled PAX2 coated control particles. There appeared to be a concentration effect with increasing PAX2 peptide 35 concentration resulting in improved Caco-2 cell binding (in the range 0.05 - 1.0 mg/g). However all absorbance readings

were low and binding of the PAX2 (5 mg/g) was not consistent with this pattern.

Using particle concentrations of 0.03125 - 0.0625 mg/well all test peptide coated particles except PAX2 (0.05 5 mg/g) exhibited comparable or greater binding to both the S100 and P100 fractions than the scrambled PAX2 coated control particles. All particles exhibited greater binding to the P100 fraction than the S100 fraction. Binding to both the S100 and P100 fractions was directly proportional to the 10 concentration of the PAX2 peptide on the particle. The absorbance readings obtained using this assay system were substantially higher than those obtained on the fixed Caco-2 cells.

The effect of blocking solution on binding of peptide-15 coated PLGA particles to P100 fractions (Caco-2 P35) was assessed using 1% bovine serum albumin (BSA) and 1% milk powder blocking solutions to assess background binding. The following particles were assayed at concentrations in the range 0.03125 - 0.0625 mg/well: no peptide; scrambled PAX2; 20 and a range of PAX2 coated particles having peptide concentrations from 5-0.05 mg/g. As previously observed using 1% BSA, all test peptide coated particles except PAX2 coated at 0.05 mg/g exhibited comparable or greater binding to the P100 fractions than the scrambled PAX2 coated control 25 particles. Binding to P100 fractions was directly proportional to the concentration of the PAX2 peptide on the particle (although in this instance PAX2 (5 mg/g) exhibited slightly lower binding than PAX2 (1 mg/g)). A similar trend was observed using 1% milk powder and a particle 30 concentration of 0.0625 mg/well. However all absorbance readings were low when 1% milk powder was used and the binding pattern was not detectable using particles at a concentration of 0.0625 mg/well.

Non-specific binding of peptide-coated PLGA particles to 35 plastic was also assessed using 1% BSA and 1% milk powder blocking solutions. The binding pattern observed above could be detected when BSA was used; however, absorbance readings

were substantially lower and binding of particles PAX2 (0.1 and 0.05 mg/g respectively) was not detectable. When 1% milk powder was used, all absorbance readings were low and no binding pattern was detectable. BSA was chosen for blocking 5 in subsequent assays.

**8.3. Comparison of Peptide-Coated Particle and Synthetic Peptide Binding to P100 fractions**

Binding of dansylated peptides to P100 fractions 10 was assessed to determine if peptide binding was predictive of peptide-coated particle binding. Figure 16 illustrates the data obtained for the dansylated peptides A) HAX42, P31 D-form and scrambled PAX2 and B) PAX2, HAX42 and scrambled PAX2.

Two consecutive assays produced substantial variations 15 in absorbance readings. Initially, the HAX42 peptide exhibited strong binding when compared to the scrambled PAX2 control. The P31 D-form peptide (ZElan053) exhibited binding at the highest dilution only. In the repeat assay, HAX42 20 also exhibited significant binding compared to the scrambled PAX2 control. However, the scrambled PAX2 control and HAX42 produced relatively high absorbance values compared to those obtained in the previous assay. The PAX2 peptide was indistinguishable from the scrambled PAX2 control. 25 Peptide/particle binding correlation is summarized as follows in Table 26:

TABLE 26

Peptide/particle assay correlation

30	Peptide	Assay correlation
	HAX42	+
	PAX2	+/-
	P31 D-form	-
	Scrambled PAX2	+/-

+ positive; +/- equivocal; - negative

35 Peptide/particle binding correlated well for the HAX42 peptide. In contrast, no correlation could be detected

for the P31 D-form (ZEelan053) peptide. Since the P31 D-form peptide-coated particles exhibited greater binding to the S100 fraction than the P100 fraction (unlike the other test peptides) it appears that the particle binding interaction 5 was non-specific or that some other molecule was competing for binding to the P100 fraction but not to the S100 fraction. Thus the peptide/particle assay correlation may be useful for distinguishing between specific and non-specific binding interactions. The scrambled PAX2 control produced 10 variable results so that it was difficult to assess the PAX2 binding correlation.

#### 8.4. Determination of HAX42 and PAX2 Binding Motif Sequences

15 Peptides and GST fusion proteins of HAX42, PAX2 and various derivatives were assayed using peptide ELISA to P100 membrane fractions derived from Caco-2 cells. The GST-PAX2 protein and PAX2 peptide data indicate that a core binding motif lies in the amino acid sequence TNAKHSSHNRRLRTR (SEQ ID NO: ) otherwise named GST-106 and ZElan033. Similarly, the 20 HAX42 peptide data suggest that a core binding motif for HAX42 lies in the amino acid sequence PGDYNCCGNCNSTG (SEQ ID NO: ), otherwise named ZElan091.

The peptides and proteins were analyzed by a 25 dansylated peptide ELISA method in which 96 well plates were coated overnight at 4°C with 100µl/well coating protein (normally 100µg/ml P100 membrane fraction) in 0.05M carbonate buffer pH9.6. Nonspecific binding was blocked using 200µl/well, 2% Marvel/PBS for 2 hours at 37°C prior to 30 incubation with dansylated peptides. The plates were washed three times with PBS/0.05% Tween 20 and after each subsequent incubation step. The peptides were diluted in blocking solution at a starting concentration of 100µg/ml and diluted 1:2 downwards, 100µl/well, followed by incubation at room 35 temperature for 1 hour, exactly. A buffer blank control was included to ensure that background binding to plastic was not due to the antibodies used in the assay system. To detect the

dansylated peptides, a mouse anti-dansyl antibody (DB3, Cytogen Corp.) at 1:1340 dilution in blocking buffer and 100 $\mu$ l/well was added followed by incubation at room temperature for 1 hour. The plates were then incubated with 5 an anti-mouse  $\lambda$ -HRP conjugated antibody (Southern Biotech 1060-05) at a 1:10,000 dilution in blocking solution, 100 $\mu$ l/well for 1 hour at room temperature. Plates were developed using 75 $\mu$ l/well Bionostics TMB substrate and incubated for approximately 10 minutes. The developing 10 reaction was stopped using Bionostics Red-Stop solution (25 $\mu$ l/well), and the optical density of the plates was read at 650nm.

GST-PAX2 Peptides - Relative Binding to P100 Fractions

15 After subtraction of the GST-peptide binding to plastic from P100 binding values, the binding of GST-PAX2 peptides were represented as a ratio of GST-HAX42 binding to P100, which was given the arbitrary value of 1.00. The following ratios were determined from binding to P100 of GST-peptides 20 at a peptide concentration of 20 $\mu$ g/ml. Bold denotes positive binding to the P100 membrane fraction.

Table 27

	GST-peptide	Value
25	<b>GST-HAX42</b>	1.00
	<b>GST-PAX2</b>	1.79
	GST-104	0.01
	GST-105	-0.08
	<b>GST-106</b>	2.71
	GST-113	0.26
	GST-114	0.17
	GST-115	0.36
30	GST	0.48

Table 28

	GST-peptide Amino Acid Sequence
GST-PAX2	<b>STPPSREAYSRPYSVDSDSDTNAKHSSHNRRLRTRSRPN</b>
GST-104	<b>STPPSREAYSRPYSVDSDSD</b>
GST-105	<b>STPPSREAYSRPYSVDSDSDTNAKHSSH</b>
5 GST-106	<b>TNAKHSSHNRRLRTRSRPN</b>
GST-113	<b>TNAKHSSH</b>
GST-114	<b>SSHNRRLRTRSRPN</b>
GST-115	<b>RRLRTRSRPN</b>

10 PAX2 Peptides - Relative Binding to P100 Fractions  
 ZElan021, full length HAX42, was given the arbitrary  
 value of 1.00 for binding to P100 at a given peptide  
 concentration determined from the signal-to-noise ratio data.  
 PAX2 and its derivatives are given as a ratio of HAX42 value  
 15 to reflect their binding abilities to P100 membrane fractions  
 derived from a Caco-2 cell line as shown in Table 29. Table  
 30 provides a line-up of the PAX2 peptides showing the  
 positive binding peptides in boldface. The GST-PAX2 peptide  
 20 and PAX2 peptide data agree, demonstrating that a binding  
 motif is in the amino acid sequence TNAKHSSHNRRLRTR (GST-106  
 and ZElan033).

25

30

35

TABLE 29

	PAX2 peptide	Binding value at 20 $\mu$ g/ml	Binding value at 20 $\mu$ g/ml	Binding value at 50 $\mu$ g/ml	Binding value at 50 $\mu$ g/ml (Jackson Ab)	Binding value at 50 $\mu$ g/ml (Southern Ab)
5	ZElan018	-0.33	1.07	0.95	1.01	
	ZElan032	1.43	2.87	0.95	1.06	
	ZElan033	0.35	1.57	0.80	0.66	
	ZElan035	0.12	0.43	0.81	0.77	
	ZElan055	0.99	0.73	1.10	0.59	
	ZElan056	0.00	0.16	0.21	0.21	
	ZElan057	0.08		0.56	0.25	
10	ZElan058	0.05		0.47	0.16	
	ZElan073	0.07		-0.11	0.49	0.66
	ZElan074	0.06		0.82	0.52	0.71
	ZElan075	0.13		0.52	0.38	0.47
	ZElan076	0.08		1.00	0.41	0.60
	ZElan077	0.20		0.76	0.54	0.73
	ZElan078	0.11		0.87	0.69	0.68
	ZElan079	0.31		0.97	0.68	0.83
	ZElan080	0.23		0.84	0.45	0.67
15	ZElan081	0.01		0.89	0.47	
	ZElan082	0.00		0.92	0.40	
	ZElan083	0.43	0.63	1.03	0.88	
	ZElan084	1.06	0.93	1.16	0.77	

20

25

30

35

Table 30

PAX2		SEQ ID NO:
Peptide	Amino acid sequence	
ZElan018	H <sub>2</sub> N-K(dns) <b>STPPSREAYSRPYSVDSDSDTNAKHSSHNRRRLTRSRPNG</b> -CONH <sub>2</sub>	
ZElan032	H <sub>2</sub> N-K(dns) <b>TNAKHSSHNRRRLTRSRPNN</b> -CONH <sub>2</sub>	
ZElan033	H <sub>2</sub> N-K(dns) <b>TNAKHSSHNRRRLTR</b> -CONH <sub>2</sub>	
5 ZElan034	H <sub>2</sub> N-K(dns) <b>SSHNRRRLTRSRPNN</b> -CONH <sub>2</sub>	
ZElan035	H <sub>2</sub> N-K(dns) <b>SSHNRRRLTR</b> -CONH <sub>2</sub>	
ZElan055	H <sub>2</sub> N-K(dns) <b>TNAKHSSHN</b> -CONH <sub>2</sub>	
ZElan056	H <sub>2</sub> N-K(dns) <b>RRLRTRSRPNN</b> -CONH <sub>2</sub>	
ZElan057	H <sub>2</sub> N-K(dns) <b>RRLRTRSR</b> -CONH <sub>2</sub>	
ZElan058	H <sub>2</sub> N-K(dns) <b>RRLRTR</b> -CONH <sub>2</sub>	
ZElan059	H <sub>2</sub> N-K(dns) <b>rrLrTrSrPN</b> -CONH <sub>2</sub>	
ZElan073	H <sub>2</sub> N-K(dns) <b>ASHNRRRLTR</b> -CONH <sub>2</sub>	
ZElan074	H <sub>2</sub> N-K(dns) <b>SAHNRRRLTR</b> -CONH <sub>2</sub>	
10 ZElan075	H <sub>2</sub> N-K(dns) <b>SSANRRRLTR</b> -CONH <sub>2</sub>	
ZElan076	H <sub>2</sub> N-K(dns) <b>SSHARRRLTR</b> -CONH <sub>2</sub>	
ZElan077	H <sub>2</sub> N-K(dns) <b>SSHNARLRLTR</b> -CONH <sub>2</sub>	
ZElan078	H <sub>2</sub> N-K(dns) <b>SSHNRALRLTR</b> -CONH <sub>2</sub>	
ZElan079	H <sub>2</sub> N-K(dns) <b>SSHNRRARTR</b> -CONH <sub>2</sub>	
ZElan080	H <sub>2</sub> N-K(dns) <b>SSHNRRRLATR</b> -CONH <sub>2</sub>	
ZElan081	H <sub>2</sub> N-K(dns) <b>SSHNRRRLRAR</b> -CONH <sub>2</sub>	
ZElan082	H <sub>2</sub> N-K(dns) <b>SSHNRRRLRTA</b> -CONH <sub>2</sub>	
SCRAMBLED PAX2 PEPTIDES:		
15 ZElan083	H <sub>2</sub> N-K(dns) <b>GRNHDVVSSNTKSYRSPRSASYPRLSNNDRTDRTEPAPSS</b> -CONH <sub>2</sub>	
ZElan084	H <sub>2</sub> N-K(dns) <b>RNTRNKTTSRLSANPHRSHR</b> -CONH <sub>2</sub>	

HAX42 Peptides - Relative Binding to P100 Fractions

ZElan021, full length HAX42, was given the arbitrary value of 1.00 for binding to P100 at a given peptide concentration determined from the signal-to-noise ratio data. HAX42 and its derivatives are given as a ratio of HAX42 value to reflect their binding abilities to P100 membrane fractions derived from a Caco-2 cell line as shown in Table 31. Table 32 provides a line-up of the HAX42 peptides showing the positive binding peptides in boldface. A core binding motif appears to lie in the amino acid sequence PGDYNCCGNCNSTG (ZElan091).

TABLE 31

	HAX42 peptide	Binding value at 20 $\mu$ g/ml	Binding value at 50 $\mu$ g/ml	Binding value at 50 $\mu$ g/ml	Binding value at 25 $\mu$ g/ml	Binding value at 25 $\mu$ g/ml	Binding value at 25 $\mu$ g/ml
5	ZElan021	1.00	1.00	1.00	1.00	1.00	1.00
	ZElan060	0.44	0.56	0.43			
	ZElan061	0.20	0.60	0.38			
	ZElan062	0.11	0.42	0.34			
	ZElan065	0.00	0.54	0.30			
	ZElan067	0.08	0.52	0.40			
	ZElan070	0.59	0.97	0.39			
	ZElan071	1.22	0.89	0.75			
	ZElan072	0.83	0.61	0.88			
	ZElan087			0.46	0.44		
	ZElan088			2.21	1.41	1.63	
	ZElan089			0.55	0.44	0.49	
10	ZElan090			2.06	1.54	2.16	
	ZElan091			2.02	1.37	1.20	
	ZElan092			1.41	1.90	0.91	
	ZElan093			1.88	1.37	1.33	

Table 32  
Amino acid sequence

	HAX42 Peptide
15	ZElan021 H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAKEPGDYNCCGNGNSTGRKVFNRRPSAIPT-CONH <sub>2</sub>
	ZElan060 H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAKEPGDYNCCGNG-CONH <sub>2</sub>
	ZElan061 H <sub>2</sub> N-K(dns) GGNNSTGRKVFNRRPSAIPT-CONH <sub>2</sub>
	ZElan062 H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAKEPG-CONH <sub>2</sub>
	ZElan065 H <sub>2</sub> N-K(dns)RKVFNRRPS-CONH <sub>2</sub>
	ZElan067 H <sub>2</sub> N-K(dns)NRRRPS-CONH <sub>2</sub>
20	ZElan070 H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAKEPGDYNCCGNGNST-CONH <sub>2</sub>
	ZElan071 H <sub>2</sub> N-K(dns)NLRSDNAKEPGDYNCCGNGNSTGRKVFN-CONH <sub>2</sub>
	ZElan072 H <sub>2</sub> N-K(dns)PGDYNCCGNGNSTGRKVFNRRPSAIPT-CONH <sub>2</sub>
	ZElan087 H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAKEPGDY-CONH <sub>2</sub>
	ZElan088 H <sub>2</sub> N-K(dns)SDNAKEPGDYNCCGNGNSTG-CONH <sub>2</sub>
	ZElan089 H <sub>2</sub> N-K(dns)SDHALGTNLRSDNAK-CONH <sub>2</sub> -CONH <sub>2</sub>
	ZElan090 H <sub>2</sub> N-K(dns)EPGDYNCCGNGNSTG
	ZElan091 H <sub>2</sub> N-K(dns)PGDYNCCGNGNSTG-CONH <sub>2</sub>
	ZElan092 H <sub>2</sub> N-K(dns)PGDYNCCGNG-CONH <sub>2</sub>
25	ZElan093 H <sub>2</sub> N-K(dns)NCCGNGNSTG-CONH <sub>2</sub>

9. FORMULATIONSGeneral Method for Preparation of Coacervated Particles.

Solid particles containing a Therapeutic as defined herein are prepared using a coacervation method. The are 30 particles are formed from a polymer and have a particle size of between about 10nm and 500  $\mu$ m, most preferably 50 to 800 nm. In addition the particles contain targeting ligands which are incorporated into the particles using a number of methods.

35 The organic phase (B) polymer of the general method given above may be soluble, permeable, impermeable,

biodegradable or gastroretentive. The polymer may consist of a mixture of polymer or copolymers and may be a natural or synthetic polymer. Representative biodegradable polymers include without limitation polyglycolides; polylactides; 5 poly(lactide-co-glycolides), including DL, L and D forms; copolyoxalates; polycaprolactone; polyesteramides; polyorthoesters; polyanhydrides; polyalkylcyanoacrylates; polyhydroxybutyrates; polyurethanes; albumin; casein; citosan derivatives; gelatin; acacia; celluloses; polysaccharides; 10 alginic acid; polypeptides; and the like, copolymers thereof, mixtures thereof and stereoisomers thereof. Representative synthetic polymers include alkyl celluloses; hydroxalkyl celluloses; cellulose ethers; cellulose esters; nitrocelluloses; polymers of acrylic and methacrylic acids 15 and esters thereof; dextrans; polyamides; polycarbonates; polyalkylenes; polyalkylene glycols; polyalkylene oxides; polyalkylene terephthalates; polyvinyl alcohols; polyvinyl ethers; polyvinyl esters; polyvinyl halides; polyvinylpyrrolidone; polysiloxanes and polyurethanes and co- 20 polymers thereof.

Typically, particles are formed using the following general method:

An aqueous solution (A) of a polymer, surface active agent, surface stabilising or modifying agent or salt, 25 or surfactant preferably a polyvinyl alcohol (PVA) or derivative with a % hydrolysis 50 - 100% and a molecular weight range 500 - 500,000, most preferably 80-100% hydrolysis and 10,000-150,000 molecular weight, is introduced into a vessel. The mixture (A) is stirred under low shear 30 conditions at 10- 2000 rpm, preferably 100-600 rpm. The pH and/or ionic strength of this solution may be modified using salts, buffers or other modifying agents. The viscosity of this solution may be modified using polymers, salts, or other viscosity enhancing or modifying agents.

35 A polymer, preferably poly(lactide-co-glycolide), polylactide, polyglycolide or a combination thereof or in any enantiomeric form or a covalent conjugate of the these

polymers with a targeting ligand is dissolved in water miscible organic solvents to form organic phase (B). Most preferably, a combination of acetone and ethanol is used in a range of ratios from 0:100 acetone: ethanol to 100: 0  
5 acetone: ethanol depending upon the polymer used.

Additional polymer(s), peptide(s) sugars, salts, natural/biological polymers or other agents may also be added to the organic phase (B) to modify the physical and chemical properties of the resultant particle product.

10 A drug or bioactive substance may be introduced into either the aqueous phase (A) or the organic phase (B). A targeting ligand may also be introduced into either the aqueous phase (A) or the organic phase (B) at this point.

15 The organic phase (B) is added into the stirred aqueous phase (A) at a continuous rate. The solvent is evaporated, preferably by a rise in temperature over ambient and/or the use of a vacuum pump. The particles are now present as a suspension (C). A targeting ligand may be introduced into the stirred suspension at this point.

20 A secondary layer of polymer(s), peptide(s) sugars, salts, natural/biological polymers or other agents may be deposited on to the pre-formed particulate core by any suitable method at this stage.

25 The particles (D) are then separated from the suspension (C) using standard colloidal separation techniques, preferably by centrifugation at high 'g' force, filtration, gel permeation chromatography, affinity chromatography or charge separation techniques. The supernatant is discarded and the particles (D) re-suspended  
30 in a washing solution (E) preferably water, salt solution, buffer or organic solvent(s). The particles (D) are separated from the washing liquid in a similar manner as previously described and re-washed, commonly twice. A targeting ligand may be dissolved in washing solution (E) at the final washing  
35 stage and may be used to wash the particles (D).

The particles may then be dried. Particles may then be further processed for example, tabletted, encapsulated or spray dried.

The release profile of the particles formed above 5 may be varied from immediate to controlled or delayed release dependent upon the formulation used and/or desired.

Drug loading may be in the range 0-90% w/w.

Targeting ligand loading may be in the range 0-90% w/w.

Specific examples include the following examples:

10

EXAMPLE 1: Peptide added at the final washing stage

Product: Bovine Insulin loaded nanoparticles

Aim: To prepare a 2g batch of insulin loaded nanoparticles at a theoretical loading of 50mg/g and with the 15 peptide ZElan018 added.

Formulation Details

RG504H	(Lot no. 250583)	2.0g
Acetone		45ml
Ethanol:		5ml
20 PVA (aq. 5%w/v)		400ml
Bovine Insulin (Lot no. 86H0674)		100mg
Peptide: PAX2 (ZElan018)		10mg/50ml dH <sub>2</sub> O

Experimental details:

25 The 5% w/v PVA solution was prepared by heating water until near boiling point, adding PVA and stirring until cool. The organic phase was prepared by adding acetone, 45ml, and ethanol, 5ml, together. The polymer solution was prepared by adding RG504H, 2g, to the organic phase and 30 stirring until dissolved. The IKA™ reactor vessel was set up, all seals greased and the temperature was set at 25°C. The PVA solution, 400ml, was added into the reactor vessel and stirred at 400 rpm.

Bovine insulin, 100mg, was added into the stirring PVA 35 solution. Using clean tubing and a green needle, the polymer solution was slowly dripped in the stirring PVA solution with the peristaltic pump set at 40. The solvent was allowed to

evaporate by opening the ports and allowing the dispersion to stir overnight at 400 rpm.

The suspension was centrifuged in a Beckman Ultracentrifuge™ with swing-out rotor at 12,500 rpm, 4°C. The 5 supernatant was decanted and discarded. The "cake" of particles was broken up and dH<sub>2</sub>O (200mls) was added to wash the particles. The centrifugation and washing steps were repeated twice.

The peptide solution, (ZElan018, 10mg in 50ml dH<sub>2</sub>O) 10 was prepared and added to the particles for a final washing stage. The suspended particles were centrifuged as before. The supernatant liquid was decanted, the 'cake' broken up, and the particles were dried in the vacuum oven.

The particles were ground, placed in a securitainer and 15 sent for analysis. The weight of particles recovered was 1.45g. A SEM showed discrete, reasonably spherical particles in the 300-500nm size range. The potency was 49.2mg/g (98.0% of label claim). Peptide loading was 2.42 µg/mg (48.4% of label claim).

20

EXAMPLE 2: Peptide added at the beginning of manufacture

**Product:** Bovine Insulin loaded nanoparticles

**Aim:** To prepare a 2g batch of insulin loaded nanoparticles at a theoretical loading of 50mg/g and with the 25 peptide ZElan018 added at the beginning of manufacture.

**Formulation Details**

RG504H	(Lot no. 250583)	2.0g
Acetone		45ml
Ethanol:		5ml
30 PVA(aq. 5%w/v)		400ml
Bovine Insulin (Lot no. 65H0640)		100mg
Peptide: PAX2 (ZElan018ii)		10mg

**Experimental details:**

35 The 5% w/v PVA solution was prepared by heating water until near boiling point, adding PVA and stirring until cool. The organic phase was prepared by adding acetone,

45ml, and ethanol, 5ml, together. The polymer solution was prepared by adding RG504H (polyactide-co-glycolide, Boehringer Ingelheim), 2g, to the organic phase prepared in step above and stirring until dissolved. The IKA™ reactor 5 vessel was set up, all seals greased and the temperature was set at 25°C. The PVA solution, 400ml, was added into the reactor vessel and stirred at 400 rpm.

Bovine insulin, 100mg, was added into the stirring PVA solution. PAX2 (ZElan018ii, 10mg) was added to the 10 stirring PVA solution. Using clean tubing and a green needle, the polymer solution was slowly dripped into the stirring PVA solution with the peristaltic pump set at 40. The solvent was allowed to evaporate by opening the ports and allowing the dispersion to stir overnight at 400 rpm. The 15 suspension was centrifuged in a Beckman Ultracentrifuge™ with swing-out rotor at 12,500 rpm, 4°C. The supernatant was decanted and discarded.

The "cake" of particles was broken up and dH<sub>2</sub>O (200ml) was added to wash the particles. The centrifugation 20 and washing steps were repeated twice. The 'cake' was broken up and the particles were dried in the vacuum oven.

The particles were ground, placed in a securitainer and sent for analysis. The weight of the particles recovered was 1.6g. The potency was 47.3mg/g (94.6% of label claim). 25 Peptide loading was 1.689μg/mg (33.8% of label claim).

#### EXAMPLE 3 Peptide added 1 hour before centrifugation

**Product:** Bovine Insulin loaded nanoparticles

**Aim:** To prepare a 1g batch of insulin loaded

30 nanoparticles at a theoretical loading of 50mg/g and with the peptide ZElan018 added 1 hour before centrifugation.

#### Formulation Details

RG504H (Lot no. 250583) 1.0g

Acetone 22.5ml

35 Ethanol: 2.5ml

PVA(aq. 5%w/v) 200ml

Bovine Insulin (Lot no. 65H0640) 50mg

Peptide: PAX2 (ZElan018) 5mg

**Experimental details:**

The 5% w/v PVA solution was prepared by heating 5 water until near boiling point, adding PVA and stirring until cool. The organic phase was prepared by adding acetone, 22.5ml, and ethanol, 2.5ml, together. The polymer solution was prepared by adding RG504H, 1g, to the organic phase prepared above and stirring until dissolved. The IKA™ 10 reactor vessel was set up, all seals greased and the temperature was set at 25°C. The PVA solution, 200ml, was added into the reactor vessel and stirred at 400 rpm.

Bovine insulin, 50mg, was added into the stirring PVA solution. Using clean tubing and a green needle, the 15 polymer solution was slowly dripped in the stirring PVA solution with the peristaltic pump set at 40. The solvent was allowed to evaporate by opening the ports and allowing the dispersion to stir overnight at 400 rpm.

PAX2 (ZElan018 5mg) was added to the stirring 20 particle suspension. After 1 hr, the suspension was centrifuged in a Beckman Ultracentrifuge™ with swing-out rotor at 12,500 rpm, 4°C. The supernatant was decanted and discarded. The "cake" of particles was broken up and dH<sub>2</sub>O (200ml) was added to wash the particles. The centrifugation 25 and washing steps were repeated twice.

The 'cake' was broken up and the particles were dried in the vacuum oven. The particles were ground, placed in a securitainer and sent for analysis. Potency was 20.75mg/g (41.5% of label claim). Peptide loading was 30 1.256μg/mg (25.12 % of label claim).

**EXAMPLE 4: Leuprolide acetate loaded nanoparticles**

Aim: To prepare a 3g batch of leuprolide-acetate loaded nanoparticles at a theoretical loading of 20mg/g and with the 35 peptide ZElan024 added.

**Formulation Details**

RG504H (Lot no. 271077) 3.0g

Acetone	67.5ml
Ethanol:	7.5ml
PVA (aq. 5%w/v)	600ml
Leuprolide acetate (Lot no. V14094)	60mg
5 Peptide: P31 (ZElan024)	15mg/50ml dH <sub>2</sub> O

**Experimental details:**

The PVA solution was prepared and the organic phase was prepared by adding acetone, 67.5ml, and ethanol, 7.5ml, 10 together. The polymer solution was prepared by adding RG504H, 3g, to the organic phase prepared above and stirring until dissolved. The IKA™ reactor vessel was set up, all seals greased and the temperature was set at 25°C. The PVA solution, 600ml, was added into the reactor vessel and 15 stirred at 400 rpm.

Leuprolide acetate, 60mg, was added into the stirring PVA solution. Using clean tubing and a green needle, the polymer solution, was slowly dripped in the stirring PVA solution with the peristaltic pump set at 40. 20 The solvent was allowed to evaporate by opening the ports and allowing the dispersion to stir overnight at 400 rpm. The suspension was centrifuged in a Beckman Ultracentrifuge™ with swing-out rotor at 15,000 rpm, 4°C. The supernatant was decanted and retained for analysis.

25 The "cake" of particles was broken up and dH<sub>2</sub>O (200ml) was added to wash the particles. The centrifugation and washing steps were repeated twice.

The peptide solution (P31 (SEQ ID NO:43), 15mg in 50ml dH<sub>2</sub>O) was prepared and added to the particles for a final 30 washing stage. The suspended particles were centrifuged as before. The supernatant liquid was decanted, and the particles were dried in the vacuum oven.

The particles were ground, placed in a securitainer and sent for analysis. The weight of particles recovered was 35 1.87g. SEM showed discrete, reasonably spherical particles in the 300-500nm size range. The potency was 4.7mg/g (23.4% of label claim). Peptide loading was 1.76μg/mg.

EXAMPLE 5: Peptide added by 'spiking' polymer phase with polymer-peptide conjugate

Product: Bovine Insulin loaded nanoparticles

Aim: To prepare a 3g batch of insulin loaded 5 nanoparticles at a theoretical loading of 50mg/g and with the polymer-peptide conjugate PLGA-ZElan019 added.

**Formulation Details**

RG504H	(Lot no. 271077)	2.85g
RG504H-ZElan019 conjugate		0.15g
10 (5PAX5-conjugate)		
Acetone		67.5ml
Ethanol:		7.5ml
PVA (aq. 5%w/v)		600ml
Bovine Insulin (Lot no. 86H0674)		150mg

15

**Experimental details:**

The 5% w/v PVA solution was prepared by heating water until near boiling point, adding PVA and stirring until cool. The organic phase was prepared by adding acetone, 20 67.5ml, and ethanol, 7.5ml, together. The polymer solution was prepared by adding RG504H and the polymer-peptide conjugate to the organic phase and stirring until dissolved.

The IKA™ reactor vessel was set up, all seals greased and the temperature was set at 25°C. The PVA 25 solution, 400ml, was added into the reactor vessel and stirred at 400 rpm.

Bovine insulin, 100mg, was added into the stirring PVA solution. Using clean tubing and a green needle, the polymer solution, was slowly dripped in the stirring PVA 30 solution with the peristaltic pump set at 40. The solvent was allowed to evaporate by opening the ports and allowing the dispersion to stir overnight at 400 rpm.

The suspension was centrifuged in a Beckman Ultracentrifuge™ with swing-out rotor at 12,500 rpm, 4°C. 35 The supernatant was decanted and discarded. The "cake" of particles was broken up and dH<sub>2</sub>O (200ml) was added to wash the

particles. The centrifugation washing step was repeated twice.

The 'cake' was broken up and the particles were dried in the vacuum oven. The particles were ground, placed 5 in a securitainer and sent for analysis. The weight of particles recovered was 2.8g. The potency was 53.1mg/g 106.2% of label claim). Peptide loading was 4.02  $\mu$ g/mg (80.4% of label claim).

10 10. ANIMAL STUDIES

Study 1

An open-loop study in which the test solution was injected directly into the ileum was done. Wistar rats (300-350g) were fasted for 4 hours and anaesthetized by 15 intramuscular administration 15 to 20 minutes prior to administration of the test solution with a solution of ketamine [0.525 ml of ketamine (100 mg/ml) and 0.875 ml of acepromazine maleate-BP ACP (2mg/ml)]. The rats were then injected with a test solution (injection volume: 1.5ml PBS) 20 intra-duodenally at 2-3 cm below the pyloris. The test solution contained either PLGA particles manufactured according to the coacervation procedure given above with or without targeting peptides or by the "spiked" method given above. Insulin (fast-acting bovine; 28.1 iu/mg) was 25 incorporated in the particles at 5% drug loading for a total of 100iu insulin (70 mg particles) or 300iu insulin (210 mg particles). Blood glucose values for the rats were measured using a Glucometer™ (Bayer; 0.1 to 33.3 m/mol/L); plasma insulin values were measured using a Phadeseph RIA Kit™ 30 (Upjohn Pharmacia; 3 to 240  $\mu$ U/ml-assayed in duplicate). Systemic and portal blood was sampled.

Study groups included animals receiving test solutions containing particles coated with the following peptides shown in Table 33.

Table 33

	Study Group I	Receptor hSI	Peptide SNi10
			SNi34
5	II	hPEPT1	P31
			5PAX5
	III	HPT1	PAX2
			HAX42
10	IV	D2H	DCX8
			DCX11
15	V ("spiked")	hPEPT1	P31-PLGA conjugate
			5PAX5-PLGA conjugate

Control groups included: 1) PBS control (1.5ml) Open-Loop; 2) Insulin solution (1iu/0.2ml) subcutaneous; 3) Insulin particles - no peptide (1iu/0.2ml) subcutaneous; 4) Insulin particles/all 8 peptides mix (1iu/0.2ml) subcutaneous; 5) Insulin loaded particles/peptide control (scrambled 5PAX5) (100iu/1.5ml) Open-Loop; 6) Insulin loaded particles/peptide control (scrambled 5PAX5) (300iu/1.5ml) Open-Loop; 7) Control particles (insulin-free)/all 8 peptide mix (equivalent 100iu/1.5ml) Open-Loop; and 8) Control particles (insulin-free)/all 8 peptide mix (equivalent 300iu/1.5ml) Open-Loop.

The following describes the pharmacokinetics for 300iu-loading:

Target Receptor	F%*	Fold-increase**	Stat. Sig.**
HPT1	10.37	17.0	<0.001
Spiked hPEPT1	4.94	7.5	0.005
PAX2 scrambled	3.50	3.6	NS
Mix-8	2.00	2.0	NS
30 hPEPT1	1.60	1.5	NS
D2H	1.57	1.4	NS
hSI	0.54	0.9	NS

\* based on area under the curve (AUC) (1-4h), base-line adjusted, relative to subcutaneous insulin solution 1iu

\*\* Fold increase in AUC compared to insulin particles: 300iu

35 Figures 17A and 17B show the systemic blood glucose and insulin levels following intestinal administration of control (PBS); insulin solution; insulin particles; all 8

peptides mix particles and study group peptide-particles (100iu). Figures 18A and 18B show the systemic blood glucose and insulin levels following intestinal administration of control (PBS); insulin solution; insulin particles and study 5 group peptide-particles (300iu).

HPT1 targeted peptide coated particles provided the most potent enhancement of the delivery of insulin over subcutaneous injection of insulin followed by hPEPT1 spiked > PAX2 scrambled > mix-8 > hPEPT1 > D2H > uncoated particles > 10 hSI > solution. In a repeat study, the uncoated particles containing insulin gave similar profiles but the HPT1-peptide targeted particles gave a reduced profile (3-fold). The insulin-free PLGA particles and the all-8 mix particles did not show an effect on the basal insulin or glucose levels. 15 The HPT1 targeting particles, the PEPT1 spiked, targeting particles, and the PEPT1 targeting particles also reduced blood glucose levels indicative that the insulin delivered was bioactive. The other targeting particles were also shown to reduce blood glucose levels although not to the same 20 extent as the HPT1 and PEPT1 spiked particles. No histological differences were observed in the small intestine for any of the formulations evaluated.

### Study 2

25 A second open-loop study, similar to study 1 above, was undertaken with the following treatment groups as shown in Table 34.

Table 34

30

Group Number	Dose Insulin (iu)	Description
1		PBS control
2a	1	subcutaneous, bovine insulin
2b	2	subcutaneous, bovine insulin
2c	3	subcutaneous, bovine insulin
2d	4	subcutaneous, bovine insulin
2e	10	subcutaneous, bovine insulin

	2f	20	subcutaneous, bovine insulin
	2g	4	subcutaneous, human insulin
	3	300	uncoated insulin particles
	4	100	HAX42/PAX2 with 300 iu particle loading
5	5	300	HAX42/PAX2 (40mer) particles
	6	300	HAX42 (40mer) particles
	7	300	HAX42 particles + 10-fold excess free HAX42 (40mer)
	8	300	PAX2 (40mer) particles
	9	300	PAX2 freeze-dried (40mer) particles
	10	300	PAX2 scrambled particles III (40mer)
10	11	300	PAX2 scrambled particles IV (19mer)
	12	300	5PAX5/P31 (40mer) particles
	13	300	P31 (40mer) particles
	14	300	5PAX5 (40mer) particles
15	15	300	HAX42 (27mer) particles
	16	300	PAX2 (20mer) particles
	17	300	P31 (20mer) particles
	18	300	PAX2 (15mer) particles
	19	300	P31 (15mer) particles
20	20	300	P31 D-form I (5 D-arginine) (16mer) particles
	21	300	P31 D-form II (2 D-arginine) (16mer) particles
	22	300	HAX42 (10mer)

Availability of insulin following administration was assessed relative to a 1 and 20iu subcutaneous dose because the response to increasing subcutaneous doses of bovine insulin does not increase linearly over the range of 1 to 20iu. Data up to three hours post-dosing was available for most animals. Therefore, availability was first assessed using individual AUC(0-3h) data estimated from baseline-subtracted data for which data up to 3 hours was available. This approach may lead to an underestimation of the availability as some animals that gave a high response often did not survive for 3 hours and, therefore, were excluded from the analyses. In an attempt to capture as much of these high responses observed at the earlier timepoints as possible, the mean baseline-subtracted plasma concentration

data was used to estimate an AUC for each group. Table 35 shows the results based on this second approach (AUC(0-3h) calculated from the mean plasma concentration data).

5

Table 35

Group	Dose iu	Mean AUC <sub>(0-3h)</sub>	F vs. 1 iu	F vs. 20 iu
1	0	2.14		
2a	1	875.27	100.00	28.86
2b	2	2439.36	139.35	40.22
10	2c	3671.44	139.82	40.36
2d	4	6912.18	197.43	56.98
2e	10	27224.41	311.04	89.77
2f	20	60651.28	346.47	100.00
2g	4	14255.49	407.17	117.52
3	300	10677.78	4.07	1.17
3 -Rat43	300	4645.06	1.77	0.51
15	4	3527.18	4.03	1.16
5	300	27112.26	10.33	2.98
6	300	33091.68	12.60	3.64
7	300	9303.09	3.54	1.02
8	300	34241.83	13.04	3.76
9	300	10968.83	4.18	1.21
10	300	27692.78	10.55	3.04
20	11	3004.29	1.14	0.33
12	300	18852.61	7.18	2.07
13	300	20278.43	7.72	2.23
14	300	17400.38	6.63	1.91
15	300	16775.69	6.39	1.84
16	300	14217.47	5.41	1.56
17	300	8197.97	3.12	0.90
25	18	25050.59	9.54	2.75
19	300	7927.96	3.02	0.87
20	300	21519.57	8.20	2.37
21	300	6322.41	2.41	0.69
22	300	12553.01	4.78	1.38

The data for group 3 (uncoated insulin particles) are  
30 expressed with and without Rat 43. This animal had an  
atypically high response to these uncoated particles and,  
therefore, may have biased the data for this group.

This data shows that a combination of peptide-coated particles (HAX42/PAX2 or 5PAX5/P31) shows no greater  
35 availability than particles coated with the individual peptides. Further, peptide-coated particles have a greater availability than uncoated peptides. Scrambling the 40mer

PAX2 peptide did not result in a loss of bioavailability. Scrambling the PAX2 peptide and reducing the size to 19mer resulted in a loss of bioavailability although this loss may be attributed in part to the reduction in peptide size.

5 Reducing peptide size resulted in loss of bioavailability. The D-form of P31 (ZElan053) had increased bioavailability possibly due to greater resistance to peptide breakdown. A competitive excess of peptide resulted in a loss of bioavailability, and freeze drying caused a loss in 10 bioavailability. By way of example, measurement of blood glucose levels showed that the HPT1 and hPEPT1 targeting particles incorporating HAX42, PAX2, P31 (SEQ ID NO:43), and P31 D-form (ZElan053) reduced blood glucose levels indicating that the insulin delivered was bioactive.

15 In further studies, insulin was recovered from the targeting particles following particle formation by dissolution and analyzed by electrophoresis in non-denaturing sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE). The analysis of the insulin by non- 20 denaturing SDS-PAGE and also by western blot transferred to membranes and subsequent screening with an antibody to insulin, indicated that the insulin was intact, with no evidence of degradation, dimerization, or aggregation during the process of particle formation.

25

### Study 3

An intraduodenal open loop model study was carried out on Wistar rats (300-350g). Group 1 was administered leuprolide acetate (12.5  $\mu$ g) subcutaneously. Group 2 was 30 administered intraduodenally uncoated leuprolide acetate particles (600  $\mu$ g, 1.5 ml). Group 3 was intraduodenally administered leuprolide acetate particles coated with PAX2 (600  $\mu$ g; 1.5 ml). Group 4 was administered intraduodenally leuprolide acetate particles coated with P31 (SEQ ID NO:43) 35 (600  $\mu$ g, 1.5 ml). Figure 19 shows the leuprolide plasma concentration following administration to these four groups. Both the P31 (SEQ ID NO:43) and the PAX2 coated leuprolide

particles administered intraduodenally provided enhanced plasma levels of leuprolide relative to subcutaneous injection.

5 Homologies of GIT transport-binding peptides to known proteins are shown in Figures 20, 21A-F, and 22 A-D.

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, 10 various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

15 Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

20

25

30

35

## SEQUENCE LISTING

## (1) GENERAL INFORMATION

(i) APPLICANTS: CYTOGEN CORPORATION and ÉLAN CORPORATION, plc

5 (ii) TITLE OF THE INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METHODS

(iii) NUMBER OF SEQUENCES: 265

## (iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Pennie & Edmonds LLP
- (B) STREET: 1155 Avenue of the Americas
- (C) CITY: New York
- (D) STATE: New York
- (E) COUNTRY: USA
- (F) ZIP: 10036

## (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette
- (B) COMPUTER: IBM Compatible
- (C) OPERATING SYSTEM: DOS
- (D) SOFTWARE: FastSEQ Version 2.0

## (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:

## (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Misrock, S. Leslie
- (B) REGISTRATION NUMBER: 18,872
- (C) REFERENCE/DOCKET NUMBER: 1101-209-228

## (ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 212-790-9090
- (B) TELEFAX: 212-869-9741
- (C) TELEX: 66141 PENNIE

25

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

30

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Arg Ser Gly Ala Tyr Glu Ser Pro Asp Gly Arg Gly Gly Arg Ser Tyr  
1 5 10 15  
Val Gly Gly Gly Gly Cys Gly Asn Ile Gly Arg Lys His Asn Leu  
20 25 30  
Trp Gly Leu Arg Thr Ala Ser Pro Ala Cys Trp Asp  
35 40

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ser Pro Arg Ser Phe Trp Pro Val Val Ser Arg His Glu Ser Phe Gly  
1 5 10 15  
Ile Ser Asn Tyr Leu Gly Cys Gly Tyr Arg Thr Cys Ile Ser Gly Thr  
20 25 30  
Met Thr Lys Ser Ser Pro Ile Tyr Pro Arg His Ser  
35 40

10 (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ser Ser Ser Ser Asp Trp Gly Gly Val Pro Gly Lys Val Val Arg Glu  
1 5 10 15  
Arg Phe Lys Gly Arg Gly Cys Gly Ile Ser Ile Thr Ser Val Leu Thr  
20 25 30  
Gly Lys Pro Asn Pro Cys Pro Glu Pro Lys Ala Ala  
35 40

20 (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

25 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Arg Val Gly Gln Cys Thr Asp Ser Asp Val Arg Arg Pro Trp Ala Arg  
1 5 10 15  
Ser Cys Ala His Gln Gly Cys Gly Ala Gly Thr Arg Asn Ser His Gly  
20 25 30  
Cys Ile Thr Arg Pro Leu Arg Gln Ala Ser Ala His  
35 40

30 (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

35 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Ser His Ser Gly Gly Met Asn Arg Ala Tyr Gly Asp Val Phe Arg Glu  
 1 5 10 15  
 Leu Arg Asp Arg Trp Asn Ala Thr Ser His His Thr Arg Pro Thr Pro  
 20 25 30  
 Gln Leu Pro Arg Gly Pro Asn  
 35

5 (2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 41 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ser Pro Cys Gly Gly Ser Trp Gly Arg Phe Met Gln Gly Gly Leu Phe  
 1 5 10 15  
 Gly Gly Arg Thr Asp Gly Cys Gly Ala His Arg Asn Arg Thr Ser Ala  
 20 25 30  
 Ser Leu Glu Pro Pro Ser Ser Asp Tyr  
 35 40

15 (2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Arg Gly Ala Ala Asp Gln Arg Arg Gly Trp Ser Glu Asn Leu Gly Leu  
 1 5 10 15  
 Pro Arg Val Gly Trp Asp Ala Ile Ala His Asn Ser Tyr Thr Phe Thr  
 20 25 30  
 Ser Arg Arg Pro Arg Pro Pro  
 35

25 (2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Ser Gly Gly Glu Val Ser Ser Trp Gly Arg Val Asn Asp Leu Cys Ala  
 1 5 10 15  
 Arg Val Ser Trp Thr Gly Cys Gly Thr Ala Arg Ser Ala Arg Thr Asp  
 20 25 30  
 35 Asn Lys Gly Phe Leu Pro Lys His Ser Ser Leu Arg  
 35 40

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Ser Asp Ser Asp Gly Asp His Tyr Gly Leu Arg Gly Gly Val Arg Cys  
 1 5 10 15  
 Ser Leu Arg Asp Arg Gly Cys Gly Leu Ala Leu Ser Thr Val His Ala  
 20 25 30  
 Gly Pro Pro Ser Phe Tyr Pro Lys Leu Ser Ser Pro  
 35 40

10

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

15

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Arg Ser Leu Gly Asn Tyr Gly Val Thr Gly Thr Val Asp Val Thr Val  
 1 5 10 15  
 Leu Pro Met Pro Gly His Ala Asn His Leu Gly Val Ser Ser Ala Ser  
 20 25 30  
 Ser Ser Asp Pro Pro Arg Arg  
 35

20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

25

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Arg Thr Thr Thr Ala Lys Gly Cys Leu Leu Gly Ser Phe Gly Val Leu  
 1 5 10 15  
 Ser Gly Cys Ser Phe Thr Pro Thr Ser Pro Pro Pro His Leu Gly Tyr  
 20 25 30  
 30 Pro Pro His Ser Val Asn  
 35

30

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ser Pro Lys Leu Ser Ser Val Gly Val Met Thr Lys Val Thr Glu Leu  
 1 5 10 15  
 Pro Thr Glu Gly Pro Asn Ala Ile Ser Ile Pro Ile Ser Ala Thr Leu  
 20 25 30  
 Gly Pro Arg Asn Pro Leu Arg  
 35

5 (2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Arg Trp Cys Gly Ala Glu Leu Cys Asn Ser Val Thr Lys Lys Phe Arg  
 1 5 10 15  
 Pro Gly Trp Arg Asp His Ala Asn Pro Ser Thr His His Arg Thr Pro  
 20 25 30  
 Pro Pro Ser Gln Ser Ser Pro  
 35

15 (2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Arg Trp Cys Gly Ala Asp Asp Pro Cys Gly Ala Ser Arg Trp Arg Gly  
 1 5 10 15  
 Gly Asn Ser Leu Phe Gly Cys Gly Leu Arg Cys Ser Ala Ala Gln Ser  
 20 25 30  
 Thr Pro Ser Gly Arg Ile His Ser Thr Ser Thr Ser  
 35 40

25 (2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Ser Lys Ser Gly Glu Gly Gly Asp Ser Ser Arg Gly Glu Thr Gly Trp  
 1 5 10 15  
 Ala Arg Val Arg Ser His Ala Met Thr Ala Gly Arg Phe Arg Trp Tyr  
 20 25 30  
 35 Asn Gln Leu Pro Ser Asp Arg  
 35

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 38 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Arg Ser Ser Ala Asn Asn Cys Glu Trp Lys Ser Asp Trp Met Arg Arg  
 1 5 10 15  
 Ala Cys Ile Ala Arg Tyr Ala Asn Ser Ser Gly Pro Ala Arg Ala Val  
 20 25 30  
 Asp Thr Lys Ala Ala Pro  
 35

10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

15

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ser Lys Trp Ser Trp Ser Ser Arg Trp Gly Ser Pro Gln Asp Lys Val  
 1 5 10 15  
 Glu Lys Thr Arg Ala Gly Cys Gly Ser Pro Ser Ser Thr Asn Cys  
 20 25 30  
 His Pro Tyr Thr Phe Ala Pro Pro Pro Gln Ala Gly  
 35 40

20

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

25

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ser Gly Phe Trp Glu Phe Ser Arg Gly Leu Trp Asp Gly Glu Asn Arg  
 1 5 10 15  
 Lys Ser Val Arg Ser Gly Cys Gly Phe Arg Gly Ser Ser Ala Gln Gly  
 20 25 30  
 30 Pro Cys Pro Val Thr Pro Ala Thr Ile Asp Lys His  
 35 40

30

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ser Glu Ser Gly Arg Cys Arg Ser Val Ser Arg Trp Met Thr Thr Trp  
 1 5 10 15  
 Gln Thr Gln Lys Gly Gly Cys Gly Ser Asn Val Ser Arg Gly Ser Pro  
 20 25 30  
 Leu Asp Pro Ser His Gln Thr Gly His Ala Thr Thr  
 35 40

5 (2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Arg Glu Trp Arg Phe Ala Gly Pro Pro Leu Asp Leu Trp Ala Gly Pro  
 1 5 10 15  
 Ser Leu Pro Ser Phe Asn Ala Ser Ser His Pro Arg Ala Leu Arg Thr  
 20 25 30  
 Tyr Trp Ser Gln Arg Pro Arg  
 35

15 (2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Arg Met Glu Asp Ile Lys Asn Ser Gly Trp Arg Asp Ser Cys Arg Trp  
 1 5 10 15  
 Gly Asp Leu Arg Pro Gly Cys Gly Ser Arg Gln Trp Tyr Pro Ser Asn  
 20 25 30  
 Met Arg Ser Ser Arg Asp Tyr Pro Ala Gly Gly His  
 35 40

25 (2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Ser His Pro Trp Tyr Arg His Trp Asn His Gly Asp Phe Ser Gly Ser  
 1 5 10 15  
 Gly Gln Ser Arg His Thr Pro Pro Glu Ser Pro His Pro Gly Arg Pro  
 20 25 30  
 35 Asn Ala Thr Ile  
 35

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Arg Tyr Lys His Asp Ile Gly Cys Asp Ala Gly Val Asp Lys Lys Ser  
 1 5 10 15  
 Ser Ser Val Arg Gly Gly Cys Gly Ala His Ser Ser Pro Pro Arg Ala  
 20 25 30  
 Gly Arg Gly Pro Arg Gly Thr Met Val Ser Arg Leu  
 35 40

10 (2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

15 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Ser Gln Gly Ser Lys Gln Cys Met Gln Tyr Arg Thr Gly Arg Leu Thr  
 1 5 10 15  
 Val Gly Ser Glu Tyr Gly Cys Gly Met Asn Pro Ala Arg His Ala Thr  
 20 25 30  
 Pro Ala Tyr Pro Ala Arg Leu Leu Pro Arg Tyr Arg  
 35 40

20 (2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

25 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Ser Gly Arg Thr Thr Ser Glu Ile Ser Gly Leu Trp Gly Trp Gly Asp  
 1 5 10 15  
 Asp Arg Ser Gly Tyr Gly Trp Gly Asn Thr Leu Arg Pro Asn Tyr Ile  
 20 25 30  
 30 Pro Tyr Arg Gln Ala Thr Asn Arg His Arg Tyr Thr  
 35 40

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

35 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Arg Trp Asn Trp Thr Val Leu Pro Ala Thr Gly Gly His Tyr Trp Thr  
 1 5 10 15  
 Arg Ser Thr Asp Tyr His Ala Ile Asn Asn His Arg Pro Ser Ile Pro  
 20 25 30  
 His Gln His Pro Thr Pro Ile  
 35

## 5 (2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Ser Trp Ser Ser Trp Asn Trp Ser Ser Lys Thr Thr Arg Leu Gly Asp  
 1 5 10 15  
 Arg Ala Thr Arg Glu Gly Cys Gly Pro Ser Gln Ser Asp Gly Cys Pro  
 20 25 30  
 Tyr Asn Gly Arg Leu Thr Thr Val Lys Pro Arg Thr  
 35 40

## 15 (2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 37 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Ser Gly Ser Leu Asn Ala Trp Gln Pro Arg Ser Trp Val Gly Gly Ala  
 1 5 10 15  
 Phe Arg Ser His Ala Asn Asn Asn Leu Asn Pro Lys Pro Thr Met Val  
 20 25 30  
 Thr Arg His Pro Thr  
 35

## 25 (2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Arg Tyr Ser Gly Leu Ser Pro Arg Asp Asn Gly Pro Ala Cys Ser Gln  
 1 5 10 15  
 Glu Ala Thr Leu Glu Gly Cys Gly Ala Gln Arg Leu Met Ser Thr Arg  
 20 25 30  
 35 Arg Lys Gly Arg Asn Ser Arg Pro Gly Trp Thr Leu  
 35 40

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Ser Val Gly Asn Asp Lys Thr Ser Arg Pro Val Ser Phe Tyr Gly Arg  
 1 5 10 15  
 Val Ser Asp Leu Trp Asn Ala Ser Leu Met Pro Lys Arg Thr Pro Ser  
 20 25 30  
 Ser Lys Arg His Asp Asp Gly  
 35

10

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 38 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

15

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Arg Trp Pro Ser Val Gly Tyr Lys Gly Asn Gly Ser Asp Thr Ile Asp  
 1 5 10 15  
 Val His Ser Asn Asp Ala Ser Thr Lys Arg Ser Leu Ile Tyr Asn His  
 20 25 30  
 Arg Arg Pro Leu Phe Pro  
 35

20

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

25

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Arg Thr Phe Glu Asn Asp Gly Leu Gly Val Gly Arg Ser Ile Gln Lys  
 1 5 10 15  
 Lys Ser Asp Arg Trp Tyr Ala Ser His Asn Ile Arg Ser His Phe Ala  
 20 25 30  
 30 Ser Met Ser Pro Ala Gly Lys  
 35

30

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Ser Tyr Cys Arg Val Lys Gly Gly Gly Glu Gly Gly His Thr Asp Ser  
 1 5 10 15  
 Asn Leu Ala Arg Ser Gly Cys Gly Lys Val Ala Arg Thr Ser Arg Leu  
 20 25 30  
 Gln His Ile Asn Pro Arg Ala Thr Pro Pro Ser Arg  
 35 40

5 (2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Ser Trp Thr Arg Trp Gly Lys His Thr His Gly Gly Phe Val Asn Lys  
 1 5 10 15  
 Ser Pro Pro Gly Lys Asn Ala Thr Ser Pro Tyr Thr Asp Ala Gln Leu  
 20 25 30  
 Pro Ser Asp Gln Gly Pro Pro  
 35

15 (2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Ser Gln Val Asp Ser Phe Arg Asn Ser Phe Arg Trp Tyr Glu Pro Ser  
 1 5 10 15  
 Arg Ala Leu Cys His Gly Cys Gly Lys Arg Asp Thr Ser Thr Thr Arg  
 20 25 30  
 Ile His Asn Ser Pro Ser Asp Ser Tyr Pro Thr Arg  
 35 40

25 (2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Ser Phe Leu Arg Phe Gln Ser Pro Arg Phe Glu Asp Tyr Ser Arg Thr  
 1 5 10 15  
 Ile Ser Arg Leu Arg Asn Ala Thr Asn Pro Ser Asn Val Ser Asp Ala  
 20 25 30  
 35 His Asn Asn Arg Ala Leu Ala  
 35

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Arg Ser Ile Thr Asp Gly Gly Ile Asn Glu Val Asp Leu Ser Ser Val  
 1 5 10 15  
 Ser Asn Val Leu Glu Asn Ala Asn Ser His Arg Ala Tyr Arg Lys His  
 20 25 30  
 Arg Pro Thr Leu Lys Arg Pro  
 35

10

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

15

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Ser Ser Lys Val Ser Ser Pro Arg Asp Pro Thr Val Pro Arg Lys Gly  
 1 5 10 15  
 Gly Asn Val Asp Tyr Gly Cys Gly His Arg Ser Ser Ala Arg Met Pro  
 20 25 30  
 Thr Ser Ala Leu Ser Ser Ile Thr Lys Cys Tyr Thr  
 35 40

20

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

25

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Arg Ala Ser Thr Gln Gly Gly Arg Gly Val Ala Pro Glu Phe Gly Ala  
 1 5 10 15  
 Ser Val Leu Gly Arg Gly Cys Gly Ser Ala Thr Tyr Tyr Thr Asn Ser  
 20 25 30  
 30 Thr Ser Cys Lys Asp Ala Met Gly His Asn Tyr Ser  
 35 40

30

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Arg Trp Cys Glu Lys His Lys Phe Thr Ala Ala Arg Cys Ser Ala Gly  
 1 5 10 15  
 Ala Gly Phe Glu Arg Asp Ala Ser Arg Pro Pro Gln Pro Ala His Arg  
 20 25 30  
 Asp Asn Thr Asn Arg Asn Ala  
 35

5 (2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Ser Phe Gln Val Tyr Pro Asp His Gly Leu Glu Arg His Ala Leu Asp  
 1 5 10 15  
 Gly Thr Gly Pro Leu Tyr Ala Met Pro Gly Arg Trp Ile Arg Ala Arg  
 20 25 30  
 Pro Gln Asn Arg Asp Arg Gln  
 35

15 (2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 38 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Ser Arg Cys Thr Asp Asn Glu Gln Cys Pro Asp Thr Gly Thr Arg Ser  
 1 5 10 15  
 Arg Ser Val Ser Asn Ala Arg Tyr Phe Ser Ser Arg Leu Leu Lys Thr  
 20 25 30  
 His Ala Pro His Arg Pro  
 35

25 (2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Ser Ala Arg Asp Ser Gly Pro Ala Glu Asp Gly Ser Arg Ala Val Arg  
 1 5 10 15  
 Leu Asn Gly Val Glu Asn Ala Asn Thr Arg Lys Ser Ser Arg Ser Asn  
 20 25 30  
 35 Pro Arg Gly Arg Arg His Pro  
 35

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 44 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Ser Ser Ala Asp Ala Glu Lys Cys Ala Gly Ser Leu Leu Trp Trp Gly  
1 5 10 15  
Arg Gln Asn Asn Ser Gly Cys Gly Ser Pro Thr Lys Lys His Leu Lys  
20 25 30  
His Arg Asn Arg Ser Gln Thr Ser Ser Ser His  
35 40

10

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

15

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Arg Pro Lys Asn Val Ala Asp Ala Tyr Ser Ser Gln Asp Gly Ala Ala  
1 5 10 15  
Ala Glu Glu Thr Ser His Ala Ser Asn Ala Ala Arg Lys Ser Pro Lys  
20 25 30  
His Lys Pro Leu Arg Arg Pro  
35

20

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

25

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Arg Gly Ser Thr Gly Thr Ala Gly Gly Glu Arg Ser Gly Val Leu Asn  
1 5 10 15  
Leu His Thr Arg Asp Asn Ala Ser Gly Ser Gly Phe Lys Pro Trp Tyr  
20 25 30  
30 Pro Ser Asn Arg Gly His Lys  
35

30

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Arg Trp Gly Trp Glu Arg Ser Pro Ser Asp Tyr Asp Ser Asp Met Asp  
 1 5 10 15  
 Leu Gly Ala Arg Arg Tyr Ala Thr Arg Thr His Arg Ala Pro Pro Arg  
 20 25 30  
 Val Leu Lys Ala Pro Leu Pro  
 35

## 5 (2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Arg Gly Trp Lys Cys Glu Gly Ser Gln Ala Ala Tyr Gly Asp Lys Asp  
 1 5 10 15  
 Ile Gly Arg Ser Arg Gly Cys Gly Ser Ile Thr Lys Asn Asn Thr Asn  
 20 25 30  
 His Ala His Pro Ser His Gly Ala Val Ala Lys Ile  
 35 40

## 15 (2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Ser Arg Glu Glu Ala Asn Trp Asp Gly Tyr Lys Arg Glu Met Ser His  
 1 5 10 15  
 Arg Ser Arg Phe Trp Asp Ala Thr His Leu Ser Arg Pro Arg Arg Pro  
 20 25 30  
 Ala Asn Ser Gly Asp Pro Asn  
 35

## 25 (2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Glu Trp Tyr Ser Trp Lys Arg Ser Ser Lys Ser Thr Gly Leu Gly Asp  
 1 5 10 15  
 Thr Ala Thr Arg Glu Gly Cys Gly Pro Ser Gln Ser Asp Gly Cys Pro  
 20 25 30  
 35 Tyr Asn Gly Arg Leu Thr Thr Val Lys Pro Arg Lys  
 35 40

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Arg Glu Phe Ala Glu Arg Arg Leu Trp Gly Cys Asp Asp Leu Ser Trp  
 1 5 10 15  
 Arg Leu Asp Ala Glu Gly Cys Gly Pro Thr Pro Ser Asn Arg Ala Val  
 20 25 30  
 Lys His Arg Lys Pro Arg Pro Arg Ser Pro Ala Leu  
 35 40

10 (2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

15 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Ser Asp His Ala Leu Gly Thr Asn Leu Arg Ser Asp Asn Ala Lys Glu  
 1 5 10 15  
 Pro Gly Asp Tyr Asn Cys Cys Gly Asn Gly Asn Ser Thr Gly Arg Lys  
 20 25 30  
 Val Phe Asn Arg Arg Arg Pro Ser Ala Ile Pro Thr  
 35 40

20 (2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

25 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Arg His Ile Ser Glu Tyr Ser Phe Ala Asn Ser His Leu Met Gly Gly  
 1 5 10 15  
 Glu Ser Lys Arg Lys Gly Cys Gly Ile Asn Gly Ser Phe Ser Pro Thr  
 20 25 30  
 30 Cys Pro Arg Ser Pro Thr Pro Ala Phe Arg Arg Thr  
 35 40

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 38 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

35 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Ser Arg Glu Ser Gly Met Trp Gly Ser Trp Trp Arg Gly His Arg Leu  
 1 5 10 15  
 Asn Ser Thr Gly Gly Asn Ala Asn Met Asn Ala Ser Leu Pro Pro Asp  
 20 25 30  
 Pro Pro Val Ser Thr Pro  
 35

5 (2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ser Thr Pro Pro Ser Arg Glu Ala Tyr Ser Arg Pro Tyr Ser Val Asp  
 1 5 10 15  
 Ser Asp Ser Asp Thr Asn Ala Lys His Ser Ser His Asn Arg Arg Leu  
 20 25 30  
 Arg Thr Arg Ser Arg Pro Asn  
 35

15 (2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

TCTCACTCCT CGAGATCCGG CGCTTATGAG AGTCCGGATG GTGGGGGGGG TCGGAGCTAT	60
GTGGGGGGCG GGGGTGGNTG TGGTAACATT GGTCGGAAGC ATAACCTGTG GGGGCTGCGT	120
ACCGCGTCGC CGGCCTGCTG GGACTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

(2) INFORMATION FOR SEQ ID NO:57:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

TCTCACTCCT CGAGTCCTCG CTCTTTCTGG CCCGTTGTGT CCCGGCATGA GTCGTTTGGG	60
ATCTCTAACT ATTTGGGNTG TGGTTATCGT ACATGTATCT CCGGCACGAT GACTAAGTCT	120
AGCCCGATTG ACCCTCGGCA TTGCGTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TCTCACTCCT CGAGTAGTAG CTCCGATTGG GGTGGTGTGC CTGGGAAGGT GGTTAGGGAG	60
CGCTTTAAGG GGCGCGGTTG TGGTATTTCC ATCACCTCCG TGCTCACTGG GAAGCCAAT	120
CCGTGTCCGG AGCCTAAGGC GGCTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

5

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TCTCACTCCT CGAGAGTTGG CCAGTGCACG GATTCTGATG TGCGGCGTCC TTGGGCCAGG	60
TCTTGCCTC ATCAGGGTTG TGGTGCAGGC ACTCGCACT CGCACGGCTG CATCACCCGT	120
CCTCTCCGCC AGGCTAGCGC TCATTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

15

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

TCTCACTCCT CGAGCCACTC CGGTGGTATG AATAGGGCCT ACAGGGATGT GTTTAGGGAG	60
CTTCGTGATC GGTGGAACGC CACTTCCAC CACACTCGCC CCACCCCTCA GCTCCCCCGT	120
GGGCCTAATT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA	162

25

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 168 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

30

TCTCACTCCT CGAGTCCGTG CGGGGGGTG TGGGGCGTT TTATGCAGGG TGGCCTTTTC	60
GGCGGTAGGA CTGATGGTTG TGGTGCCTCAT AGAAACCGCA CTTCTGCCTC GTTAGAGCCC	120
CCGAGCAGCG ACTACTCTAG AATCGAAGGT CGCGCTAGAC TTGAGA	168

35

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

TCTCACTCCT CGAGGGGGCGC CGCCGATCAG CGGCGGGGGT GGTCCGAGAA CTTGGGGTTG	60
CCTAGGGTGG GGTGGGACGC CATCGCTCAC AATAGCTATA CGTTCACCTC GCGCCGCCCG	120
CGCCCCCCCCT CTAGA	135

## (2) INFORMATION FOR SEQ ID NO:63:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

TCTCACTCCT CGAGCGGTGG GGAGGTCAGC TCCTGGGGCC GCGTGAATGA CCTCTGCGCT	60
AGGGTGAGTT GGACTGGTTG TGGTACTGCT CGTCCGCGC GTACCGACAA CAAAGGCTTT	120
CTTCCTAAGC ACTCGTCACT CCGCTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

## (2) INFORMATION FOR SEQ ID NO:64:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

20

TCTCACTCCT CGAGTGATAG TGACGGGGAT CATTATGGGC TTCCGGGGGG GGTGC GTTGT	60
TCGCTCGTG ATAGGGGTG TGGTCTGGCC CTGTCCACCG TCCATGCTGG TCCCCCTCT	120
TTTTACCCCA AGCTCTCCAG CCCCTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

## (2) INFORMATION FOR SEQ ID NO:65:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

30

TCTCACTCCT CGAGGAGCTT GGGTAATTAT GGCGTCACCG GGACTGTGGA CGTGACGGTT	60
TTGCCCATGC CTGGCCACGC CAACCACCTT GGTGTCTCCT CCGCCTCTAG CTCTGATCCT	120
CCGCGGCCT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA	162

## (2) INFORMATION FOR SEQ ID NO:66:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 159 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

TCTCACTCCT CGAGAACTAC GACGGCTAAG GGGTGTCTTC TCGGAAGCTT CGGCGTTCTT 60  
 AGTGGGTGCT CATTACGCC AACCTCTCCA CCGCCCCACC TAGGATAACCC CCCCCACTCC 120  
 GTCAATTCTA GAATCGAAGG TCGCGCTAGA CCTTCGAGA 159

## (2) INFORMATION FOR SEQ ID NO:67:

## (i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 162 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

10 TCTCACTCCT CGAGCCGAA GTTGTCCAGC GTGGGTGTTA TGACTAAGGT CACGGAGCTG 60  
 CCCACGGAGG GGCCTAACGC CATTAGTATT CCGATCTCCG CGACCCCTCGG CCCGCGAAC 120  
 CCGCTCCGCT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA 162

## (2) INFORMATION FOR SEQ ID NO:68:

## (i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 162 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TCTCACTCCT CGAGGTGGTG CGGCCTGAG CTGTGCAACT CGGTGACTAA GAAGTTTCGC 60  
 CGGGGCTGGC GGGATCACGC CAATCCCTCC ACCCATCATC GTACTCCCCC GCCCAGCCAG 120  
 20 TCCAGCCCTT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA 162

## (2) INFORMATION FOR SEQ ID NO:69:

## (i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 176 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TCTCACTCCT CGAGGTGGTG CGGCCTGAT GACCCGTGTG GTGCCAGTCG TTGGCGGGGG 60  
 GGCAACAGCT TGTGGTTG TGGTCTTCGT TGTAGTCGG CGCAGAGCAC CCCGAGTGGC 120  
 AGGATCCATT CCACCTCGAC CAGCTCTAGA ATCGAAGGTG CGCTAGACCT TCGAGA 176

## (2) INFORMATION FOR SEQ ID NO:70:

## (i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 162 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

TCTCACTCCT CGAGTAAGTC CGGGGAGGGG GGTGACAGTA GCAGGGCGA GACGGGCTGG 60  
 GCGAGGGTTC GGTCTCACGC CATGACTGCT GGCGCTTTC GGTGGTACAA CCAGTTGCC 120

TCTGATCGGT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA

162

## (2) INFORMATION FOR SEQ ID NO:71:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 159 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

TCTCACTCCT CGAGGTGAG CGCCAATAAT TCGGAGTGG AGTCTGATTG GATGCGCAGG	60
GCCTGTATTG CTCGTTACGC CAACAGTTCG GGCCCCGCC GCGCCGTCGA CACTAAGGCC	120
10 GCGCCCTCTA GAATCGAAGG TCGCGCTAGA CCTTCGAGA	159

10

## (2) INFORMATION FOR SEQ ID NO:72:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TCTCACTCCT CGAGTAAGTG GTCGTGGAGT TCGAGGTGGG GCTCCCCGCA GGATAAGGTT	60
GAGAAGACCA GGGCGGGTTG TGGTGGTAGT CCCAGCAGCA CCAATTGTCA CCCCTACACC	120
TTTGCCCCCCC CCCCCGCAAGC CGGCTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

20

## (2) INFORMATION FOR SEQ ID NO:73:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

TCTCACTCCT CGAGTGGGTT CTGGGAGTTT AGCAGGGGGC TTTGGGATGG GGAGAACCGT	60
AAGAGTGTCC GGTGGGGTTG TGGTTTTCGT GGCTCCTCTG CTCAGGGCCC GTGTCCGGTC	120
ACGCCTGCCA CCATTGACAA ACACTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

30

## (2) INFORMATION FOR SEQ ID NO:74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

TCTCACTCCT CGAGTGAGAG CGGGCGGTGC CGTAGCGTGA GCCGGTGGAT GACGACGTGG	60
CAGACGCAGA AGGGCGGGTTG TGGTTCCAAT GTTCCCCCG GTCGCCCCCT CGACCCCTCT	120
CACCAAGACCG GGCATGCCAC TACTTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

## (2) INFORMATION FOR SEQ ID NO:75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

TCTCACTCCT	CGAGGGAGTG	GAGGTTGCC	GGGCCGCCGT	TGGACCTGTG	GGCGGGTCCG	60
AGCTTGCCCT	CTTTAACGC	CAGTTCCCAC	CCTCGGCC	TGCGCACCTA	TTGGTCCCAG	120
CGGCCCCGCT	CTAGAACATCGA	AGGTCGCGCT	AGACCTTCGA	GA		162

10

## (2) INFORMATION FOR SEQ ID NO:76:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

TCTCACTCCT	CGAGGGATGGA	GGACATCAAG	AACTCGGGGT	GGAGGGACTC	TTGTAGGTGG	60
GGTAGACCTGA	GGCCTGGTTG	TGGTAGCCGC	CAGTGGTACC	CCTCGAATAT	GCGTTCTAGC	120
AGAGATTAC	CCGCAGGGGGG	CCACTCTAGA	ATCGAAGGTC	GCGCTAGACC	TTCGAGA	177

## (2) INFORMATION FOR SEQ ID NO:77:

20

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 152 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

25

TCTCACTCCT	CGAGTCATCC	GTGGTACAGG	CATTGGAACC	ATGGTGACTT	CTCTGGTTCG	60
GGCCAGTCAC	GCCACACCCC	GCCGGAGAGC	CCCCACCCCG	GCCGCCCTAA	TGCCACCATT	120
TCTAGAACATCG	AAGGTCGCGC	TAGACCTTCG	AG			152

## (2) INFORMATION FOR SEQ ID NO:78:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

35

TCTCACTCCT	CGAGATATAA	GCACGATATC	GGTTGCGATG	CTGGGGTTGA	CAAGAACGTCG	60
TCGTCTGTGC	GTGGTGGTTG	TGGTGCTCAT	TNGTCGCCAC	CCCGCGCCGG	CCGTGGTCCT	120
CGCGGCACGA	TGGTTAGCAG	GCTTTCTAGA	ATCGAAGGTC	GCGCTAGACC	TTCGAGA	177

## (2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 177 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

TCTCACTCCT CGAGTCAGGG CTCCAAGCAG TGTATGCAGT ACCGCACCGG TCGTTTGACG	60
GTGGGGTCTG AGTATGGTTG TGGTATGAAC CCCGCCGCC ATGCCACGCC CGCTTATCCG	120
GCGCGCCTGC TGCCACGCTA TCGCTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

(2) INFORMATION FOR SEQ ID NO:80:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 177 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

TCTCACTCCT CGAGTGGCG GACTACTAGT GAGATTTCTG GGCTCTGGGG TTGGGGTGAC	60
GACCGGAGCG GTTATGGTTG GGGTAACACG CTCCGCCCA ACTACATCCC TTATAGGCAG	120
GCGACGAACA GGCATCGTTA TACGTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

(2) INFORMATION FOR SEQ ID NO:81:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 162 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

TCTCACTCCT CGAGGTGGAA TTGGACTGTC TTGCCGCCA CTGGCGGCCA TTACTGGACG	60
CGTTCGACGG ACTATCACGC CATTAAACAAT CACAGGCCGA GCATCCCCA CCAGCATCCG	120
ACCCCTATCT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA	162

(2) INFORMATION FOR SEQ ID NO:82:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 177 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

TCTCACTCCT CGAGTTGGTC GTCGTGAAT TGGAGCTCTA AGACTACTCG TCTGGCGAC	60
AGGGCGACTC GGGAGGGTTG TGGTCCCAGC CAGTCTGATG GCTGTCCTTA TAACGGCCGC	120
CTTACGACCG TCAAGCCTCG CACGTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

35 (2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 156 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

5 TCTCACTCCT CGAGTGGTAG TTTGAACGCA TGGCAACCGC GGTCAATGGGT GGGGGCGCG 60  
 TTCCGGTCAC ACGCCAACAA TAACTTGAAC CCCAAGCCCA CCATGGTTAC TNGTCACCT 120  
 ACCTCTAGAA TCGAAGGTCG CGCTAGACCT TCGAGA 156

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 178 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

10 TCTCACTCCT CGAGGTATTG GGGTTTGTCC CCGCGGGACA ACGGTCCCAG TTGTAGTCAG 60  
 15 GAGGCTACCT TGGAGGGTTG TGGTGCAGAG AGGCTGATGT CCACCCGTCG CAAGGGCGC 120  
 AACTCCGCC CGGGTGGAC GCTCTCTAGA ATCGAAGGTC GCGCTAGACC CTTCGAGA 178

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

25 TCTCACTCCT CGAGCGTGGG GAATGATAAG ACTAGCAGGC CGGTTTCCTT CTACGGCGC 60  
 GTTAGTGATC TGTGGAACGC CAGCTTGATG CCGAAGCGTA CTCCAGCTC GAAGCGCCAC 120  
 GATGATGGCT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA 162

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

TCTCACTCCT CGAGTACTCC CCCCAGTAGG GAGGCGTATA GTAGGCCCTA TAGTGTGAT 60  
 AGCGATTGCG ATACGAACGC CAAGCACAGC TCCCAACACC GCCGTNTGCG GACGCGCAGC 120  
 CGCCCGAACT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA 162

(2) INFORMATION FOR SEQ ID NO:87:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 159 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

5 TCTCACTCCT CGAGATGGCC TAGTGTGGGT TACAAGGGTA ATGGCAGTGA CACTATTGAT 60  
 GTTCACAGCA ATGACGCCAG TACTAAGAGG TCCCTCATCT ATAACCACCG CCGCCCCNTC 120  
 TTTCCCTCTA GAATCGAAGG TCGCGCTAGA CCTTCGAGA 159

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

TCTCACTCCT CGAGAACGTT TGAGAACGAC GGGCTGGGCG TCGGCCGGTC TATTCAAGAAG 60  
 AAGTCGGATA GGTGGTACGC CAGCCACAAAC ATTCGTAGCC ATTCGCGTC CATGTCTCCC 120  
 GCTGGTAAGT CTAGAACGAG AGGTCGCGCT AGACCTTCGA GA 162

15

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 160 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

TCTCACTCCT CGAGCTATTG TCGGGTTAAG GGTGGTGGGG AGGGGGGGCA TACGGATTCC 60  
 AATCTGGCTA GGTGGGTTG TGGTAAGGTG GCCAGGACCA GCAGGCTTCA GCATATCAAC 120  
 CCGCGCGCTA CCCCCCCCCTC CCGGTCTAGA ATCGAAGGTC 160

(2) INFORMATION FOR SEQ ID NO:90:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

TCTCACTCCT CGAGTTGGAC TCGGTGGGGC AAGCACANTC ATGGGGGGTT TGTGAACAAG 60  
 TCTCCCCCTG GGAAGAACGC CACGAGCCCC TACACCGACG CCCAGCTGCC CAGTGATCAG 120  
 GGTCTCCCT CTAGAACGAG AGGTCGCGCT AGACCTTCGA GA 162

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

TCTCACTCCT CGAGTCAGGT TGATTCGTTT CGTAATAGCT TTCCGGTGGTA TGAGCCGAGC	60
AGGGCTCTGT GCCATGGTTG TGGTAAGCGC GACACCTCCA CCACTCGTAT CCACAATAGC	120
CCCAGCGACT CCTATCCTAC ACGCTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

5

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

TCTCACTCCT CGAGCTTTT GCGGTTCCAG AGTCCGAGGT TCGAGGATT A CAGTAGGACG	60
ATCTNTCGGT TCGCGAACGC CACGAACCCG AGTAATGTCT CCGATGCGCA CAATAACCGG	120
GCCTTGGCCT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA	162

15

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

TCTCACTCCT CGAGGGACAT CACCGACGGG GGCATCAATG AGGTGGACCT GAGTAGTGTG	60
TCGAACGTTT TTGAGAACGC CAACTCGCAT AGGGCCTACA GGAAGCATCG CCCGACCTTG	120
AAGCGTCCTT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA	162

25

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

30

TCTCACTCCT CGAGTTCGAA GGTGAGCAGC CCGAGGGATC CGACGGTCCC GCGGAAGGGC	60
GGCAATGTTG ATTATGGTTG TGGTCACAGG TCTTCCGCC CGATGCCTAC CTCCGCTCTG	120
TCGTCGATCA CGAAGTGCTA CACTTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

TCTCACTCCT CGAGAGCCAG TANGCAGGGC GGCCGGGGTG TTGCCCCCTGA GTTTGGGGCG	60
AGCGTTTGG GTNGTGGTTG TGGTAGCGCC ACTTATTACA CGAACTCCAC CAGCTGCAAG	120
GATGCTATGG GCCACAACTA CTCGTCTAGA ATCGAAGGTC GCGNTAGACC TTCGAGA	177

## (2) INFORMATION FOR SEQ ID NO:96:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

TCTCACTCCT CGAGATGGTG CGAGAACGAC AAGTTTACGG CTGCGCGTTG CAGCGCGGG	60
GGGGGTTTTG AGAGGGANGC CAGCCGTCCG CCCCAGCCTG CCCACCGGGA TAATACCAAC	120
CGTAATGCNT NTAGAATCGA AGGTCGCGCT AGACCTTCGA GA	162

## (2) INFORMATION FOR SEQ ID NO:97:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

20

TCTCACTCCT CGAGTTTTCA GGTGTACCCG GACCATGGTC TGGAGAGGCA TGCTTTGGAC	60
GGGACGGGTC CGCTTTACGC CATGCCCGGC CGCTGGATTA GGGCGCGTCC GCAGAACAGG	120
GACGCCAGT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA	162

## (2) INFORMATION FOR SEQ ID NO:98:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 159 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

30

TCTCACTCCT CGAGCAGGTG TACGGACAAC GAGCAGTGCC CCGATACCGG GANTAGGTCT	60
CGTCCGTCA GTAACGCCAG GTACTTTTCG AGCAGGTTGC TCAAGACTCA CGCCCCCAT	120
CGCCCTCTA GAATCGAAGG TCGCGCTAGA CCTTCGAGA	159

## (2) INFORMATION FOR SEQ ID NO:99:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

TCTCACTCCT CGAGTGCCAG GGATAGCGGG CCTGC GGAGG ATGGGTCCCG CGCCGTCCGG 60  
 TTGAACGGGG TTGAGAACGC CAACACTAGG AAGTCCTCCC GCAGTAACCC GCGGGGTAGG 120  
 CGCCATCCCT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA 162

## (2) INFORMATION FOR SEQ ID NO:100:

## (i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 177 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

10 TCTCACTCCT CGAGTTCCGC CGATGCGGAG AAGTGTGCGG GCAGTCTGTT GTGGTGGGGT 60  
 AGGCAGAACAA ACTCCGGTTG TGGTCGCC ACAGAAGAAC ATCTGAAGCA CCGCAATCGC 120  
 AGTCAGACCT CCTCTTCGTC CCACCTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA 177

## (2) INFORMATION FOR SEQ ID NO:101:

## (i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 162 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

20 TCTCACTCCT CGAGACCGAA GAACGTGGCC GATGCTTATT CGTCTCAGGA CGGGGCGGCG 60  
 GCCGAGGAGA CGTCTCACGC CAGTAATGCC GCGCGGAAGT CCCCTAAGCA CAAGCCCTTG 120  
 AGGCGGCCTT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA 162

## (2) INFORMATION FOR SEQ ID NO:102:

## (i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 162 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

TCTCACTCCT CGAGAGGCAG TACGGGGACG GCCGGCGGCG AGCGTTCCGG GGTGCTAAC 60  
 CTGCACACCA GGGATAACGC CAGCGGCAGC GGTTTCAAAC CGTGGTACCC TTCAATCGG 120  
 GGTCAACAAGT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA 162

## 30 (2) INFORMATION FOR SEQ ID NO:103:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 162 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

TCTCACTCCT CGAGGTGGGG GTGGGAGAGG AGTCCGTCCG ACTACGATTG TGATATGGAC 60  
 TTGGGGCGA GGAGGTACGC CACCGCACC CACCGCGC CGCCCTCGCGT CTTGAAGGCT 120

CCCCCTGCCCT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA

162

## (2) INFORMATION FOR SEQ ID NO:104:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

TCTCACTCCT CGAGGCACGT GAAAGTGCAG GGGCTCTCAGG CTGCCTACGG GGACAAGGAT	60
ATCGGGAGGT CCAGGGGTTG TGGTTCCATT ACAAAAGAATA ACACATAATCA CGCCCATCCT	120
10 AGCCACGGCG CCGTTGCTAA GATCTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

## (2) INFORMATION FOR SEQ ID NO:105:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

TCTCACTCCT CGAGCCGCGA GGAGGCGAAC TGGGACGGCT ATAAGAGGGA GATGAGCCAC	60
CGGAGTCGCT TTTGGGACGC CACCCACCTG TCCCGCCCTC GCCGCCCGC TAACTCTGGT	120
15 GACCCTAACT CTAGAATCGA AGGTCGCGCT AGACCTTCGA GA	162

20

## (2) INFORMATION FOR SEQ ID NO:106:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

TCTCACTCNT CGAGAGAGTT CGCGGAGAGG AGGTTGTGGG GGTGTGATGA CCTGAGTTGG	60
CGTCTCGACG CGGAGGGTTG TGGTCCCCT CCGAGCAATC GGGCCGTCAA GCATCGCAAG	120
15 CCCCCGCCAC GCTCCCCCGC ACTCTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

## (2) INFORMATION FOR SEQ ID NO:107:

30

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

TCTCACTCNT NGAGTGATCA CGCGTTGGGG ACGAATCTGA GGTCTGACAA TGCCAAGGAG	60
CGGGGTGATT ACAACTGTTG TGGTAAACGGG AACTCTACCG GGCAGAAAGGT TTTTAACCGT	120
15 AGGCGCCCT CCGCCATCCC CANTTCTAGA ATCGAAGGTC GCGCTAGACC TTGAGA	177

## (2) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

TCTCACTCCT CGAGGCATAT TTCTGAGTAT AGCTTTGCGA ATTCCCACCTT GATGGGTGGC	60
GAGTCCAAGC GGAAGGGTTG TGGTATTAAC GGCTCCTTT CTCCCACTTG TCCCCGCTCC	120
CCCACCCAG CCTTCCGCCG CACCTCTAGA ATCGAAGGTC GCGCTAGACC TTCGAGA	177

10

## (2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 158 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

TCTCACTCCT CGAGCCGGGA GAGCGGGATG TGGGGTAGTT GGTGGCGTGG TCACAGGTTG	60
AATTCCACGG GGGGTAACGC CAACATGAAT GCTAGTCTGC CCCCCGACCC CCCTGTTCC	120
ACTCCGTCTA GAATCGAAGG TCGCGCTAGA CCTTCGAG	158

## (2) INFORMATION FOR SEQ ID NO:110:

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 708 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

25

Met	Gly	Met	Ser	Lys	Ser	His	Ser	Phe	Phe	Gly	Tyr	Pro	Leu	Ser	Ile		
1																15	
Phe		Phe		Ile		Val		Val		Asn		Glu		Phe		Cys	10
																25	
																30	
Met		Arg		Ala		Ile		Leu		Ile		Tyr		Phe		Thr	35
																40	
																45	
Asp		Asn		Leu		Ser		Thr		Ala		Ile		Tyr		50	
																55	
																60	
Leu		Thr		Pro		Ile		Leu		Gly		Ala		Leu		Cys	65
																70	
																75	
Phe		Lys		Thr		Ile		Val		Ser		Ile		Val		Tyr	85
																90	
																95	
Val		Thr		Ser		Val		Ser		Ile		Asn		Asp		His	100
																105	
																110	
Gly		Thr		Pro		Asp		Ser		Leu		Pro		Val		His	115
																120	
																125	
Leu		Ala		Leu		Ile		Ala		Gly		Thr		Gly		Ile	130
																135	
																140	
Ala		Phe		Gly		Gly		Asp		Gln		Phe		Glu		Gly	145
																150	
																155	
Arg		Phe		Phe		Ser		Ile		Phe		Tyr		Leu		Ala	165
																170	
																175	

Ser Thr Ile Ile Thr Pro Met Leu Arg Val Gln Gln Cys Gly Ile His  
 180 185 190  
 Ser Lys Gln Ala Cys Tyr Pro Leu Ala Phe Gly Val Pro Ala Ala Leu  
 195 200 205  
 Met Ala Val Ala Leu Ile Val Phe Val Leu Gly Ser Gly Met Tyr Lys  
 210 215 220  
 Lys Phe Lys Pro Gln Gly Asn Ile Met Gly Lys Val Ala Lys Cys Ile  
 5 225 230 235 240  
 Gly Phe Ala Ile Lys Asn Arg Phe Arg His Arg Ser Lys Ala Phe Pro  
 245 250 255  
 Lys Arg Glu His Trp Leu Asp Trp Ala Lys Glu Lys Tyr Asp Glu Arg  
 260 265 270  
 Leu Ile Ser Gln Ile Lys Met Val Thr Arg Val Met Phe Leu Tyr Ile  
 275 280 285  
 Pro Leu Pro Met Phe Trp Ala Leu Phe Asp Gln Gln Gly Ser Arg Trp  
 290 295 300  
 10 Thr Leu Gln Ala Thr Thr Met Ser Gly Lys Ile Gly Ala Leu Glu Ile  
 305 310 315 320  
 Gln Pro Asp Gln Met Gln Thr Val Asn Ala Ile Leu Ile Val Ile Met  
 325 330 335  
 Val Pro Ile Phe Asp Ala Val Leu Tyr Pro Leu Ile Ala Lys Cys Gly  
 340 345 350  
 Phe Asn Phe Thr Ser Leu Lys Lys Met Ala Val Gly Met Val Leu Ala  
 355 360 365  
 Ser Met Ala Phe Val Val Ala Ala Ile Val Gln Val Glu Ile Asp Lys  
 15 370 375 380  
 Thr Leu Pro Val Phe Pro Lys Gly Asn Glu Val Gln Ile Lys Val Leu  
 385 390 395 400  
 Asn Ile Gly Asn Asn Thr Met Asn Ile Ser Leu Pro Gly Glu Met Val  
 405 410 415  
 Thr Leu Gly Pro Met Ser Gln Thr Asn Ala Phe Met Thr Phe Asp Val  
 420 425 430  
 Asn Lys Leu Thr Arg Ile Asn Ile Ser Ser Pro Gly Ser Pro Val Thr  
 435 440 445  
 20 Ala Val Thr Asp Asp Phe Lys Gln Gly Gln Arg His Thr Leu Leu Val  
 450 455 460  
 Trp Ala Pro Asn His Tyr Gln Val Val Lys Asp Gly Leu Asn Gln Lys  
 465 470 475 480  
 Pro Glu Lys Gly Glu Asn Gly Ile Arg Phe Val Asn Thr Phe Asn Glu  
 485 490 495  
 Leu Ile Thr Ile Thr Met Ser Gly Lys Val Tyr Ala Asn Ile Ser Ser  
 500 505 510  
 Tyr Asn Ala Ser Thr Tyr Gln Phe Phe Pro Ser Gly Ile Lys Gly Phe  
 25 515 520 525  
 Thr Ile Ser Ser Thr Glu Ile Pro Pro Gln Cys Gln Pro Asn Phe Asn  
 530 535 540  
 Thr Phe Tyr Leu Glu Phe Gly Ser Ala Tyr Thr Tyr Ile Val Gln Arg  
 545 550 555 560  
 Lys Asn Asp Ser Cys Pro Glu Val Lys Val Phe Glu Asp Ile Ser Ala  
 565 570 575  
 Asn Thr Val Asn Met Ala Leu Gln Ile Pro Gln Tyr Phe Leu Leu Thr  
 580 585 590  
 30 Cys Gly Glu Val Val Phe Ser Val Thr Gly Leu Glu Phe Ser Tyr Ser  
 595 600 605  
 Gln Ala Pro Ser Asn Met Lys Ser Val Leu Gln Ala Gly Trp Leu Leu  
 610 615 620  
 Thr Val Ala Val Gly Asn Ile Ile Val Leu Ile Val Ala Gly Ala Gly  
 625 630 635 640  
 Gln Phe Ser Lys Gln Trp Ala Glu Tyr Ile Leu Phe Ala Ala Leu Leu  
 645 650 655  
 Leu Val Val Cys Val Val Phe Ala Ile Met Ala Arg Phe Tyr Thr Tyr  
 35 660 665 670  
 Ile Asn Pro Ala Glu Ile Glu Ala Gln Phe Asp Glu Asp Glu Lys Lys  
 675 680 685  
 Asn Arg Leu Glu Lys Ser Asn Pro Tyr Phe Met Ser Gly Ala Asn Ser  
 690 695 700

Gln Lys Gln Met  
705

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

TCCGGACTCT CATAAGCGCC GG

22

10

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

ACAAACGGGCC AGAAAGAGCG AG

22

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

25

ACACCACCCC AATCGGAGCT AC

22

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

TCAGAAATCCG TGCAC TGGCC AA

22

(2) INFORMATION FOR SEQ ID NO:115:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GCCCTATTCA TACCACCGGA GT

22

(2) INFORMATION FOR SEQ ID NO:116:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

CATCAGTCCT ACCGCCGAAA AG

22

(2) INFORMATION FOR SEQ ID NO:117:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

CGTATAGCTA TTGTGAGCGA TG

22

20

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

ACGCGCGGAA CGAGCAGTAC CA

22

(2) INFORMATION FOR SEQ ID NO:119:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

CCATAATGAT CCCCGTCACT AT

22

35

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

5 AGACACCCCT TAGCCGTCGT AG

22

(2) INFORMATION FOR SEQ ID NO:121:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

AGCTCCGTGA CCTTAGTCAT AA

22

(2) INFORMATION FOR SEQ ID NO:122:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

TGCACAGCTC AGCGCCGCAC CA

22

(2) INFORMATION FOR SEQ ID NO:123:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

ACGGGTCATC AGCGCCGCAC CA

22

30 (2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

TGTCACCCCT CTCCCCGGAC TT

22

## (2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

ACTCGCAATT ATTGGCGCTC GA

22

## (2) INFORMATION FOR SEQ ID NO:126:

10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

15

GTCTTCTCAA CCTTATCCTG CG

22

## (2) INFORMATION FOR SEQ ID NO:127:

20

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

AAAGCCCCCT GCTAAACTCC CA

22

## (2) INFORMATION FOR SEQ ID NO:128:

25

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

CTGCGTCTGC CACGTCGTCA TC

22

## (2) INFORMATION FOR SEQ ID NO:129:

35

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

GTTAAAAGAG GGCAAGCTCG GA

22

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 22 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

10 CCGAGTTCTT GATGTCCTCC AT

22

(2) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

TCCAATGCCT GTACCACGGA TG

22

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

25 TCGCAACCGA TATCGTGCTT AT

22

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

TGCATACACT GCTTGGAGCC CT

22

(2) INFORMATION FOR SEQ ID NO:134:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

GAAATCTCAC TAGTAGTCCG CC

22

5

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

GCGGGCAAGA CAGTCCAATT CC

22

15

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

20

GAGCTCCAAT TCCACGACGA CC

22

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

GGTTGCCATG CGTTCAAACT AC

22

(2) INFORMATION FOR SEQ ID NO:138:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

35

TCCCGCGGGG ACAAACCCGA AT

22

(2) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

CTGCTAGTCT TATCATTCCC CA

22

(2) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

CTATCGACAC TATAGGGCCT AC

22

15 (2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

TACCCTTGTA ACCCACACTA GG

22

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

30 TTCTTCTGAA TAGACCGGCC GA

22

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

CCACCAACCT TAACCCGACA AT

22

## (2) INFORMATION FOR SEQ ID NO:144:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

AGGGGGAGAC TTGTTCACAA AC

22

10

## (2) INFORMATION FOR SEQ ID NO:145:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

CGGCTCATAAC CACCGAAAGC TA

22

## (2) INFORMATION FOR SEQ ID NO:146:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

ATCGTCCTAC TGTAATCCTC GA

22

25

## (2) INFORMATION FOR SEQ ID NO:147:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

GACACACTAC TCAGGTCCAC CT

22

## (2) INFORMATION FOR SEQ ID NO:148:

35

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

CCATAATCAA CATTGCCGCC CT

22

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

10 CAAAACGCTC GCCCCAAACT CA

22

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

GTAAACTTGT GCTTCTCGCA CC

22

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

25 CCATGGTCCG GGTACACCTG AA

22

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GTTACTAACG GAACGAGACC TA

22

(2) INFORMATION FOR SEQ ID NO:153:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

TGTTGGCGTT CTCAACCCCG TT

22

5

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

ACAAACCGGAG TTGTTCTGCC TA

22

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

20

TAAGCATCGG CCACGTTCTT CG

22

(2) INFORMATION FOR SEQ ID NO:156:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

TTATCCCTGG TGTGCAGGTT GA

22

(2) INFORMATION FOR SEQ ID NO:157:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

35

TATCAGAACATC GTAGTCGGAC GG

22

(2) INFORMATION FOR SEQ ID NO:158:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

CTTTGTAATG GAACCACAAC CC

22

(2) INFORMATION FOR SEQ ID NO:159:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

CGGTGGCTCA TCTCCCTCTT AT

22

15 (2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

ATCAGACTGG CTGGGACCAAC AA

22

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

30 CACAAACCTCC TCTCCGCGAA CT

22

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

AGATTCGTCC CCAACGCGTG AT

22

## (2) INFORMATION FOR SEQ ID NO:163:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

GGGAATTCGC AAAAGCTATAAC TC

22

10

## (2) INFORMATION FOR SEQ ID NO:164:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

CCCCGTGGAA TTCAACCTGT GA

22

## (2) INFORMATION FOR SEQ ID NO:165:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

GTCGTCTTTC CAGACGT

17

25

## (2) INFORMATION FOR SEQ ID NO:166:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

CTTGCATGCC TGCAGGTGCA C

21

## (2) INFORMATION FOR SEQ ID NO:167:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

Arg Ile Ala Gly Leu Pro Trp Tyr Arg Cys Arg Thr Val Ala Phe Glu  
 1 5 10 15  
 Thr Gly Met Gln Asn Thr Gln Leu Cys Ser Thr Ile Val Gln Leu Ser  
 5 20 25 30  
 Phe Thr Pro Glu Glu  
 35

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- 10 (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

Arg Glu Phe Ala Glu Arg Arg Leu Trp Gly Cys Asp Asp Leu Ser Trp  
 1 5 10 15  
 15 Arg Leu Asp Ala Glu Gly Cys Gly Pro Thr Pro Ser Asn Arg Ala Val  
 20 25 30  
 Lys His Arg Lys Pro Arg Pro Arg Ser Pro Ala Leu  
 35 40

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
- (B) TYPE: amino acid
- 20 (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

Ser Gly Ser His Ser Gly Gly Met Asn Arg Ala Tyr Gly Asp Val Phe  
 25 1 5 10 15  
 Arg Glu Leu Arg Asp Arg Trp Tyr Ala Thr Ser His His Thr Arg Pro  
 20 25 30  
 Thr Pro Gln Leu Pro Arg Gly Pro Asn  
 35 40

(2) INFORMATION FOR SEQ ID NO:170:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

35 Ser Thr Pro Pro Ser Arg Glu Ala Tyr Ser Arg Pro Tyr Ser Val Asp  
 1 5 10 15  
 Ser Asp Ser Asp  
 20

## (2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

5

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

Ser Thr Pro Pro Ser Arg Glu Ala Tyr Ser Arg Pro Tyr Ser Val Asp  
1 5 10 15  
Ser Asp Ser Asp Thr Asn Ala Lys His Ser Ser His Asn  
20 25

10

## (2) INFORMATION FOR SEQ ID NO:172:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

15

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

Thr Asn Ala Lys His Ser Ser His Asn Arg Arg Leu Arg Thr Arg Ser  
1 5 10 15  
Arg Pro Asn

20

## (2) INFORMATION FOR SEQ ID NO:173:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

Thr Asn Ala Lys His Ser Ser His Asn  
1 5

## (2) INFORMATION FOR SEQ ID NO:174:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

35

Ser Ser His Asn Arg Arg Leu Arg Thr Arg Ser Arg Pro Asn  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:175:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

Arg Arg Leu Arg Thr Arg Ser Arg Pro Asn  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:176:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 708 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

15 Met Gly Met Ser Lys Ser His Ser Phe Phe Gly Tyr Pro Leu Ser Ile  
 1 5 10 15  
 Phe Phe Ile Val Val Asn Glu Phe Cys Glu Arg Phe Ser Tyr Tyr Gly  
 20 25 30  
 Met Arg Ala Ile Leu Ile Leu Tyr Phe Thr Asn Phe Ile Ser Trp Asp  
 35 40 45  
 Asp Asn Leu Ser Thr Ala Ile Tyr His Thr Phe Val Ala Leu Cys Tyr  
 50 55 60  
 Leu Thr Pro Ile Leu Gly Ala Leu Ile Ala Asp Ser Trp Leu Gly Lys  
 65 70 75 80  
 20 Phe Lys Thr Ile Val Ser Leu Ser Ile Val Tyr Thr Ile Gly Gln Ala  
 85 90 95  
 Val Thr Ser Val Ser Ser Ile Asn Asp Leu Thr Asp His Asn His Asp  
 100 105 110  
 Gly Thr Pro Asp Ser Leu Pro Val His Val Val Leu Ser Leu Ile Gly  
 115 120 125  
 Leu Ala Leu Ile Ala Leu Gly Thr Gly Gly Ile Lys Pro Cys Val Ser  
 130 135 140  
 25 Ala Phe Gly Gly Asp Gln Phe Glu Glu Gly Gln Glu Lys Gln Arg Asn  
 145 150 155 160  
 Arg Phe Phe Ser Ile Phe Tyr Leu Ala Ile Asn Ala Gly Ser Leu Leu  
 165 170 175  
 Ser Thr Ile Ile Thr Pro Met Leu Arg Val Gln Gln Cys Gly Ile His  
 180 185 190  
 Ser Lys Gln Ala Cys Tyr Pro Leu Ala Phe Gly Val Pro Ala Ala Leu  
 195 200 205  
 Met Ala Val Ala Leu Ile Val Phe Val Leu Gly Ser Gly Met Tyr Lys  
 210 215 220  
 30 Lys Phe Lys Pro Gln Gly Asn Ile Met Gly Lys Val Ala Lys Cys Ile  
 225 230 235 240  
 Gly Phe Ala Ile Lys Asn Arg Phe Arg His Arg Ser Lys Ala Phe Pro  
 245 250 255  
 Lys Arg Glu His Trp Leu Asp Trp Ala Lys Glu Lys Tyr Asp Glu Arg  
 260 265 270  
 Leu Ile Ser Gln Ile Lys Met Val Thr Arg Val Met Phe Leu Tyr Ile  
 275 280 285  
 35 Pro Leu Pro Met Phe Trp Ala Leu Phe Asp Gln Gln Gly Ser Arg Trp  
 290 295 300  
 Thr Leu Gln Ala Thr Thr Met Ser Gly Lys Ile Gly Ala Leu Glu Ile  
 305 310 315 320  
 Gln Pro Asp Gln Met Gln Thr Val Asn Ala Ile Leu Ile Val Ile Met

	325	330	335	
	Val Pro Ile Phe Asp Ala Val Leu Tyr	Pro Leu Ile Ala Lys	Cys Gly	
	340	345	350	
	Phe Asn Phe Thr Ser Leu Lys Lys	Met Ala Val Gly	Met Val Leu Ala	
	355	360	365	
	Ser Met Ala Phe Val Val Ala Ala Ile Val Gln Val	Glu Ile Asp Lys		
	370	375	380	
5	Thr Leu Pro Val Phe Pro Lys Gly Asn Glu Val	Gln Ile Lys Val Leu		
	385	390	395	400
	Asn Ile Gly Asn Asn Thr Met Asn Ile Ser Leu Pro	Gly Glu Met Val		
	405	410	415	
	Thr Leu Gly Pro Met Ser Gln Thr Asn Ala Phe Met	Thr Phe Asp Val		
	420	425	430	
	Asn Lys Leu Thr Arg Ile Asn Ile Ser Ser Pro	Gly Ser Pro Val Thr		
	435	440	445	
10	Ala Val Thr Asp Asp Phe Lys Gln Gly Gln Arg	His Thr Leu Leu Val		
	450	455	460	
	Trp Ala Pro Asn His Tyr Gln Val Val Lys Asp	Gly Leu Asn Gln Lys		
	465	470	475	480
	Pro Glu Lys Gly Glu Asn Gly Ile Arg Phe Val Asn	Thr Phe Asn Glu		
	485	490	495	
	Leu Ile Thr Ile Met Ser Gly Lys Val Tyr Ala Asn	Ile Ser Ser		
	500	505	510	
	Tyr Asn Ala Ser Thr Tyr Gln Phe Phe Pro Ser	Gly Ile Lys Gly Phe		
	515	520	525	
15	Thr Ile Ser Ser Thr Glu Ile Pro Pro Gln Cys	Gln Pro Asn Phe Asn		
	530	535	540	
	Thr Phe Tyr Leu Glu Phe Gly Ser Ala Tyr	Thr Tyr Ile Val Gln Arg		
	545	550	555	560
	Lys Asn Asp Ser Cys Pro Glu Val Lys Val	Phe Glu Asp Ile Ser Ala		
	565	570	575	
	Asn Thr Val Asn Met Ala Leu Gln Ile Pro Gln	Tyr Phe Leu Leu Thr		
	580	585	590	
20	Cys Gly Glu Val Val Phe Ser Val Thr Gly Leu	Glu Phe Ser Tyr Ser		
	595	600	605	
	Gln Ala Pro Ser Asn Met Lys Ser Val Leu Gln	Ala Gly Trp Leu Leu		
	610	615	620	
	Thr Val Ala Val Gly Asn Ile Ile Val Leu Ile	Val Ala Gly Ala Gly		
	625	630	635	640
	Gln Phe Ser Lys Gln Trp Ala Glu Tyr Ile Leu	Phe Ala Ala Leu Leu		
	645	650	655	
	Leu Val Val Cys Val Val Phe Ala Ile Met Ala	Arg Phe Tyr Thr Tyr		
	660	665	670	
25	Ile Asn Pro Ala Glu Ile Glu Ala Gln Phe Asp	Glu Asp Glu Lys Lys		
	675	680	685	
	Asn Arg Leu Glu Lys Ser Asn Pro Tyr Phe Met	Ser Gly Ala Asn Ser		
	690	695	700	
	Gln Lys Gln Met			
	705			

## (2) INFORMATION FOR SEQ ID NO:177:

## 30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3345 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

## 35 (A) NAME/KEY: Coding Sequence

- (B) LOCATION: 88...2583

- (D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

GAATTCCGTC	TCGACCAC	TG AATGGAAGAA	AAGGACTTTT	AACCACCATT	TTGTGACTTA	60
CAGAAAGGAA	TTTGAATAAA	GAAA	ACT ATG ATA CTT	CAG GCC CAT CTT	CAC TCC	114
			Met Ile Leu Gln Ala	His Leu His Ser		
			1	5		
5	CTG TGT CTT CTT ATG CTT TAT TTG GCA ACT GGA TAT GGC CAA GAG GGG					162
	Leu Cys Leu Leu Met Leu Tyr Leu Ala Thr Gly Tyr Gly Gln Glu Gly					
10	10 15	20	25			
	AAG TTT AGT GGA CCC CTG AAA CCC ATG ACA TTT TCT ATT TAT GAA GGC					210
	Lys Phe Ser Gly Pro Leu Lys Pro Met Thr Phe Ser Ile Tyr Glu Gly					
	30 35	40				
10	CAA GAA CCG AGT CAA ATT ATA TTC CAG TTT AAG GCC AAT CCT CCT GCT					258
	Gln Glu Pro Ser Gln Ile Ile Phe Gln Phe Lys Ala Asn Pro Pro Ala					
	45 50	55				
	GTG ACT TTT GAA CTA ACT GGG GAG ACA GAC AAC ATA TTT GTG ATA GAA					306
	Val Thr Phe Glu Leu Thr Gly Glu Thr Asp Asn Ile Phe Val Ile Glu					
	60 65	70				
15	CGG GAG GGA CTT CTG TAT TAC AAC AGA GCC TTG GAC AGG GAA ACA AGA					354
	Arg Glu Gly Leu Leu Tyr Tyr Asn Arg Ala Leu Asp Arg Glu Thr Arg					
	75 80	85				
	TCT ACT CAC AAT CTC CAG GTT GCA GCC CTG GAC GCT AAT GGA ATT ATA					402
	Ser Thr His Asn Leu Gln Val Ala Ala Leu Asp Ala Asn Gly Ile Ile					
	90 95	100	105			
	GTG GAG GGT CCA GTC CCT ATC ACC ATA GAA GTG AAG GAC ATC AAC GAC					450
	Val Glu Gly Pro Val Pro Ile Thr Ile Glu Val Lys Asp Ile Asn Asp					
	110 115	120				
20	AAT CGA CCC ACG TTT CTC CAG TCA AAG TAC GAA GGC TCA GTA AGG CAG					498
	Asn Arg Pro Thr Phe Leu Gln Ser Lys Tyr Glu Gly Ser Val Arg Gln					
	125 130	135				
	AAC TCT CGC CCA GGA AAG CCC TTC TTG TAT GTC AAT GCC ACA GAC CTG					546
	Asn Ser Arg Pro Gly Lys Pro Phe Leu Tyr Val Asn Ala Thr Asp Leu					
	140 145	150				
25	GAT GAT CCG GCC ACT CCC AAT GGC CAG CTT TAT TAC CAG ATT GTC ATC					594
	Asp Asp Pro Ala Thr Pro Asn Gly Gln Leu Tyr Tyr Gln Ile Val Ile					
	155 160	165				
	CAG CTT CCC ATG ATC AAC AAT GTC ATG TAC TTT CAG ATC AAC AAC AAA					642
	Gln Leu Pro Met Ile Asn Asn Val Met Tyr Phe Gln Ile Asn Asn Lys					
	170 175	180	185			
30	ACG GGA GCC ATC TCT CTT ACC CGA GAG GGA TCT CAG GAA TTG AAT CCT					690
	Thr Gly Ala Ile Ser Leu Thr Arg Glu Gly Ser Gln Glu Leu Asn Pro					
	190 195	200				
	GCT AAG AAT CCT TCC TAT AAT CTG GTG ATC TCA GTG AAG GAC ATG GGA					738
	Ala Lys Asn Pro Ser Tyr Asn Leu Val Ile Ser Val Lys Asp Met Gly					
	205 210	215				
35	GGC CAG AGT GAG AAT TCC TTC AGT GAT ACC ACA TCT GTG GAT ATC ATA					786
	Gly Gln Ser Glu Asn Ser Phe Ser Asp Thr Thr Ser Val Asp Ile Ile					
	220 225	230				
	G TG A CA G AG A AT T T GG A AA G C A C C A A A C C T G T G G AG AT G G T G G A A					834
	Val Thr Glu Asn Ile Trp Lys Ala Pro Lys Pro Val Glu Met Val Glu					
	235 240	245				

	AAC TCA ACT GAT CCT CAC CCC ATC AAA ATC ACT CAG GTG CGG TGG AAT	882	
	Asn Ser Thr Asp Pro His Pro Ile Lys Ile Thr Gln Val Arg Trp Asn		
250	255	260	265
	GAT CCC GGT GCA CAA TAT TCC TTA GTT GAC AAA GAG AAG CTG CCA AGA	930	
	Asp Pro Gly Ala Gln Tyr Ser Leu Val Asp Lys Glu Lys Leu Pro Arg		
270	275	275	280
5	TTC CCA TTT TCA ATT GAC CAG GAA GAT ATT TAC GTG ACT CAG CCC	978	
	Phe Pro Phe Ser Ile Asp Gln Glu Gly Asp Ile Tyr Val Thr Gln Pro		
285	290	295	
	TTG GAC CGA GAA GAA AAG GAT GCA TAT GTT TTT TAT GCA GTT GCA AAG	1026	
	Leu Asp Arg Glu Glu Lys Asp Ala Tyr Val Phe Tyr Ala Val Ala Lys		
300	305	310	
10	GAT GAG TAC GGA AAA CCA CTT TCA TAT CCG CTG GAA ATT CAT GTA AAA	1074	
	Asp Glu Tyr Gly Lys Pro Leu Ser Tyr Pro Leu Glu Ile His Val Lys		
315	320	325	
	GTT AAA GAT ATT AAT GAT AAT CCA CCT ACA TGT CCG TCA CCA GTA ACC	1122	
	Val Lys Asp Ile Asn Asp Asn Pro Pro Thr Cys Pro Ser Pro Val Thr		
330	335	340	345
15	GTA TTT GAG GTC CAG GAG AAT GAA CGA CTG GGT AAC AGT ATC GGG ACC	1170	
	Val Phe Glu Val Gln Glu Asn Glu Arg Leu Gly Asn Ser Ile Gly Thr		
350	355	360	
	CTT ACT GCA CAT GAC AGG GAT GAA GAA AAT ACT GCC AAC AGT TTT CTA	1218	
	Leu Thr Ala His Asp Arg Asp Glu Glu Asn Thr Ala Asn Ser Phe Leu		
365	370	375	
	AAC TAC AGG ATT GTG GAG CAA ACT CCC AAA CTT CCC ATG GAT GGA CTC	1266	
	Asn Tyr Arg Ile Val Glu Gln Thr Pro Lys Leu Pro Met Asp Gly Leu		
20	380	385	390
	TTC CTA ATC CAA ACC TAT GCT GGA ATG TTA CAG TTA GCT AAA CAG TCC	1314	
	Phe Leu Ile Gln Thr Tyr Ala Gly Met Leu Gln Leu Ala Lys Gln Ser		
395	400	405	
	TTG AAG AAG CAA GAT ACT CCT CAG TAC AAC TTA ACG ATA GAG GTG TCT	1362	
	Leu Lys Lys Gln Asp Thr Pro Gln Tyr Asn Leu Thr Ile Glu Val Ser		
410	415	420	425
25	GAC AAA GAT TTC AAG ACC CTT TGT TTT GTG CAA ATC AAC GTT ATT GAT	1410	
	Asp Lys Asp Phe Lys Thr Leu Cys Phe Val Gln Ile Asn Val Ile Asp		
430	435	440	
	ATC AAT GAT CAG ATC CCC ATC TTT GAA AAA TCA GAT TAT GGA AAC CTG	1458	
	Ile Asn Asp Gln Ile Pro Ile Phe Glu Lys Ser Asp Tyr Gly Asn Leu		
445	450	455	
30	ACT CTT GCT GAA GAC ACA AAC ATT GGG TCC ACC ATC TTA ACC ATC CAG	1506	
	Thr Leu Ala Glu Asp Thr Asn Ile Gly Ser Thr Ile Leu Thr Ile Gln		
460	465	470	
	GCC ACT GAT GCT GAT GAG CCA TTT ACT GGG AGT TCT AAA ATT CTG TAT	1554	
	Ala Thr Asp Ala Asp Glu Pro Phe Thr Gly Ser Ser Lys Ile Leu Tyr		
475	480	485	
35	CAT ATC ATA AAG GGA GAC AGT GAG GGA CGC CTG GGG GTT GAC ACA GAT	1602	
	His Ile Ile Lys Gly Asp Ser Glu Gly Arg Leu Gly Val Asp Thr Asp		
490	495	500	505
	CCC CAT ACC AAC ACC GGA TAT GTC ATA ATT AAA AAG CCT CTT GAT TTT	1650	
	Pro His Thr Asn Thr Gly Tyr Val Ile Ile Lys Lys Pro Leu Asp Phe		

	510	515	520	
	GAA ACA GCA GCT GTT TCC AAC ATT GTG TTC AAA GCA GAA AAT CCT GAG Glu Thr Ala Ala Val Ser Asn Ile Val Phe Lys Ala Glu Asn Pro Glu 525	530	535	1698
5	CCT CTA GTG TTT GGT GTG AAG TAC AAT GCA AGT TCT TTT GCC AAG TTC Pro Leu Val Phe Gly Val Lys Tyr Asn Ala Ser Ser Phe Ala Lys Phe 540	545	550	1746
	ACG CTT ATT GTG ACA GAT GTG AAT GAA GCA CCT CAA TTT TCC CAA CAC Thr Leu Ile Val Thr Asp Val Asn Glu Ala Pro Gln Phe Ser Gln His 555	560	565	1794
10	GTA TTC CAA GCG AAA GTC AGT GAG GAT GTA GCT ATA GGC ACT AAA GTG Val Phe Gln Ala Lys Val Ser Glu Asp Val Ala Ile Gly Thr Lys Val 570	575	580	1842
	GGC AAT GTG ACT GCC AAG GAT CCA GAA GGT CTG GAC ATA AGC TAT TCA Gly Asn Val Thr Ala Lys Asp Pro Glu Gly Leu Asp Ile Ser Tyr Ser 590	595	600	1890
	CTG AGG GGA GAC ACA AGA GGT TGG CTT AAA ATT GAC CAC GTG ACT GGT Leu Arg Gly Asp Thr Arg Gly Trp Leu Lys Ile Asp His Val Thr Gly 605	610	615	1938
15	GAG ATC TTT AGT GTG GCT CCA TTG GAC AGA GAA GCC GGA AGT CCA TAT Glu Ile Phe Ser Val Ala Pro Leu Asp Arg Glu Ala Gly Ser Pro Tyr 620	625	630	1986
	CGG GTA CAA GTG GTG GCC ACA GAA GTA GGG GGG TCT TCC TTA AGC TCT Arg Val Gln Val Val Ala Thr Glu Val Gly Ser Ser Leu Ser Ser 635	640	645	2034
20	GTG TCA GAG TTC CAC CTG ATC CTT ATG GAT GTG AAT GAC AAC CCT CCC Val Ser Glu Phe His Leu Ile Leu Met Asp Val Asn Asp Asn Pro Pro 650	655	660	2082
	AGG CTA GCC AAG GAC TAC ACG GGC TTG TTC TGC CAT CCC CTC AGT Arg Leu Ala Lys Asp Tyr Thr Gly Leu Phe Phe Cys His Pro Leu Ser 670	675	680	2130
25	GCA CCT GGA AGT CTC ATT TTC GAG GCT ACT GAT GAT GAT CAG CAC TTA Ala Pro Gly Ser Leu Ile Phe Glu Ala Thr Asp Asp Asp Gln His Leu 685	690	695	2178
	TTT CGG GGT CCC CAT TTT ACA TTT TCC CTC GGC AGT GGA AGC TTA CAA Phe Arg Gly Pro His Phe Thr Phe Ser Leu Gly Ser Gly Ser Leu Gln 700	705	710	2226
30	AAC GAC TGG GAA GTT TCC AAA ATC AAT GGT ACT CAT GCC CGA CTG TCT Asn Asp Trp Glu Val Ser Lys Ile Asn Gly Thr His Ala Arg Leu Ser 715	720	725	2274
	ACC AGG CAC ACA GAC TTT GAG GAG AGG GCG TAT GTC GTC TTG ATC CGC Thr Arg His Thr Asp Phe Glu Glu Arg Ala Tyr Val Val Leu Ile Arg 730	735	740	2322
	ATC AAT GAT GGG GGT CGG CCA CCC TTG GAA GGC ATT GTT TCT TTA CCA Ile Asn Asp Gly Gly Arg Pro Pro Leu Glu Gly Ile Val Ser Leu Pro 750	755	760	2370
35	GTT ACA TTC TGC AGT TGT GTG GAA GGA AGT TGT TTC CGG CCA GCA GGT Val Thr Phe Cys Ser Cys Val Glu Gly Ser Cys Phe Arg Pro Ala Gly 765	770	775	2418

	CAC CAG ACT GGG ATA CCC ACT GTG GGC ATG GCA GTT GGT ATA CTG CTG His Gln Thr Gly Ile Pro Thr Val Gly Met Ala Val Gly Ile Leu Leu 780 785 790	2466
	ACC ACC CTT CTG GTG ATT GGT ATA ATT TTA GCA GTT GTG TTT ATC CGC Thr Thr Leu Leu Val Ile Gly Ile Ile Leu Ala Val Val Phe Ile Arg 795 800 805	2514
5	ATA AAG AAG GAT AAA GGC AAA GAT AAT GTT GAA AGT GCT CAA GCA TCT Ile Lys Lys Asp Lys Gly Lys Asp Asn Val Glu Ser Ala Gln Ala Ser 810 815 820 825	2562
	GAA GTC AAA CCT CTG AGA AGC TGAATTTGAA AAGGAATGTT TGAATTTATA TAGC Glu Val Lys Pro Leu Arg Ser 830	2617
10	AAGTGCTATT TCAGCAACAA CCATCTCATC CTATTACTTT TCATCTAACG TGCATTATAA TTTTTTAAAC AGATATTCCC TCTTGTCCCT TAATATTGCT TAAATATTTC TTTTTGAGG TGGAGTCTTG CTCTGTGCC CAGGCTGGAG TACAGTGGTG TGATCCCAGC TCACTGCAAC CTCCGCCTCC TGGGTTCACA TGATTCTCCT GCCTCAGCTT CCTAAGTAGC TGGGTTTACA GGCACCCACC ACCATGCCCA GCTAATTTT GTATTTTAA TAGAGACGGG GTTTCGCCAT TTGGGCCAGGC TGGTCTTGAA CTCCTGACGT CAAGTGATCT GGCTGCCTTG GTCTCCCAAT ACAGGCATGA ACCACTGCAC CCACCTACTT AGATATTCA TGTCATAG ACATTAGAGA GATTTTTCAT TTTTCATGA CATTTCCTCT CTCTGCAAAT GGCTTAGCTA CTTGTGTTT TCCCTTTGG GGCAAGACAG ACTCATTAAA TATTCTGTAC ATTTTTCTT TATCAAGGAG 15 ATATATCAGT GTTGCTCAT AGAACTGCT GGATTCCATT TATGTTTTT CTGATTCCAT CCTGTGTCCC CTTCATCCTT GACTCCTTGT GTATTCACT GAATTTCAAA CATTGTCAG AGAAGAAAAA AGTGAGGACT CAGGAAAAAT AAATAAATAA AAGAACAGCC TTTTGCAGGC GCGAATTC	2677 2737 2797 2857 2917 2977 3037 3097 3157 3217 3277 3337 3345

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 832 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

25	Met Ile Leu Gln Ala His Leu His Ser Leu Cys Leu Leu Met Leu Tyr 1 5 10 15
	Leu Ala Thr Gly Tyr Gly Gln Glu Gly Lys Phe Ser Gly Pro Leu Lys 20 25 30
	Pro Met Thr Phe Ser Ile Tyr Glu Gly Gln Glu Pro Ser Gln Ile Ile 35 40 45
	Phe Gln Phe Lys Ala Asn Pro Pro Ala Val Thr Phe Glu Leu Thr Gly 50 55 60
	Glu Thr Asp Asn Ile Phe Val Ile Glu Arg Glu Gly Leu Leu Tyr Tyr 65 70 75 80
30	Asn Arg Ala Leu Asp Arg Glu Thr Arg Ser Thr His Asn Leu Gln Val 85 90 95
	Ala Ala Leu Asp Ala Asn Gly Ile Ile Val Glu Gly Pro Val Pro Ile 100 105 110
	Thr Ile Glu Val Lys Asp Ile Asn Asp Asn Arg Pro Thr Phe Leu Gln 115 120 125
	Ser Lys Tyr Glu Gly Ser Val Arg Gln Asn Ser Arg Pro Gly Lys Pro 130 135 140
35	Phe Leu Tyr Val Asn Ala Thr Asp Leu Asp Asp Pro Ala Thr Pro Asn 145 150 155 160
	Gly Gln Leu Tyr Tyr Gln Ile Val Ile Gln Leu Pro Met Ile Asn Asn 165 170 175
	Val Met Tyr Phe Gln Ile Asn Asn Lys Thr Gly Ala Ile Ser Leu Thr 180 185 190

Arg Glu Gly Ser Gln Glu Leu Asn Pro Ala Lys Asn Pro Ser Tyr Asn  
 195 200 205  
 Leu Val Ile Ser Val Lys Asp Met Gly Gly Gln Ser Glu Asn Ser Phe  
 210 215 220  
 Ser Asp Thr Thr Ser Val Asp Ile Ile Val Thr Glu Asn Ile Trp Lys  
 225 230 235 240  
 Ala Pro Lys Pro Val Glu Met Val Glu Asn Ser Thr Asp Pro His Pro  
 245 250 255  
 5 Ile Lys Ile Thr Gln Val Arg Trp Asn Asp Pro Gly Ala Gln Tyr Ser  
 260 265 270  
 Leu Val Asp Lys Glu Lys Leu Pro Arg Phe Pro Phe Ser Ile Asp Gln  
 275 280 285  
 Glu Gly Asp Ile Tyr Val Thr Gln Pro Leu Asp Arg Glu Glu Lys Asp  
 290 295 300  
 Ala Tyr Val Phe Tyr Ala Val Ala Lys Asp Glu Tyr Gly Lys Pro Leu  
 305 310 315 320  
 10 Ser Tyr Pro Leu Glu Ile His Val Lys Val Lys Asp Ile Asn Asp Asn  
 325 330 335  
 Pro Pro Thr Cys Pro Ser Pro Val Thr Val Phe Glu Val Gln Glu Asn  
 340 345 350  
 Glu Arg Leu Gly Asn Ser Ile Gly Thr Leu Thr Ala His Asp Arg Asp  
 355 360 365  
 Glu Glu Asn Thr Ala Asn Ser Phe Leu Asn Tyr Arg Ile Val Glu Gln  
 370 375 380  
 Thr Pro Lys Leu Pro Met Asp Gly Leu Phe Leu Ile Gln Thr Tyr Ala  
 385 390 395 400  
 15 Gly Met Leu Gln Leu Ala Lys Gln Ser Leu Lys Lys Gln Asp Thr Pro  
 405 410 415  
 Gln Tyr Asn Leu Thr Ile Glu Val Ser Asp Lys Asp Phe Lys Thr Leu  
 420 425 430  
 Cys Phe Val Gln Ile Asn Val Ile Asp Ile Asn Asp Gln Ile Pro Ile  
 435 440 445  
 Phe Glu Lys Ser Asp Tyr Gly Asn Leu Thr Leu Ala Glu Asp Thr Asn  
 450 455 460  
 20 Ile Gly Ser Thr Ile Leu Thr Ile Gln Ala Thr Asp Ala Asp Glu Pro  
 465 470 475 480  
 Phe Thr Gly Ser Ser Lys Ile Leu Tyr His Ile Ile Lys Gly Asp Ser  
 485 490 495  
 Glu Gly Arg Leu Gly Val Asp Thr Asp Pro His Thr Asn Thr Gly Tyr  
 500 505 510  
 Val Ile Ile Lys Lys Pro Leu Asp Phe Glu Thr Ala Ala Val Ser Asn  
 515 520 525  
 Ile Val Phe Lys Ala Glu Asn Pro Glu Pro Leu Val Phe Gly Val Lys  
 530 535 540  
 25 Tyr Asn Ala Ser Ser Phe Ala Lys Phe Thr Leu Ile Val Thr Asp Val  
 545 550 555 560  
 Asn Glu Ala Pro Gln Phe Ser Gln His Val Phe Gln Ala Lys Val Ser  
 565 570 575  
 Glu Asp Val Ala Ile Gly Thr Lys Val Gly Asn Val Thr Ala Lys Asp  
 580 585 590  
 Pro Glu Gly Leu Asp Ile Ser Tyr Ser Leu Arg Gly Asp Thr Arg Gly  
 595 600 605  
 30 Trp Leu Lys Ile Asp His Val Thr Gly Glu Ile Phe Ser Val Ala Pro  
 610 615 620  
 Leu Asp Arg Glu Ala Gly Ser Pro Tyr Arg Val Gln Val Val Ala Thr  
 625 630 635 640  
 Glu Val Gly Gly Ser Ser Leu Ser Ser Val Ser Glu Phe His Leu Ile  
 645 650 655  
 Leu Met Asp Val Asn Asp Asn Pro Pro Arg Leu Ala Lys Asp Tyr Thr  
 660 665 670  
 Gly Leu Phe Phe Cys His Pro Leu Ser Ala Pro Gly Ser Leu Ile Phe  
 675 680 685  
 35 Glu Ala Thr Asp Asp Asp Gln His Leu Phe Arg Gly Pro His Phe Thr  
 690 695 700  
 Phe Ser Leu Gly Ser Gly Ser Leu Gln Asn Asp Trp Glu Val Ser Lys  
 705 710 715 720

Ile Asn Gly Thr His Ala Arg Leu Ser Thr Arg His Thr Asp Phe Glu  
 725 730 735  
 Glu Arg Ala Tyr Val Val Leu Ile Arg Ile Asn Asp Gly Gly Arg Pro  
 740 745 750  
 Pro Leu Glu Gly Ile Val Ser Leu Pro Val Thr Phe Cys Ser Cys Val  
 755 760 765  
 Glu Gly Ser Cys Phe Arg Pro Ala Gly His Gln Thr Gly Ile Pro Thr  
 770 775 780  
 5 Val Gly Met Ala Val Gly Ile Leu Leu Thr Thr Leu Leu Val Ile Gly  
 785 790 795 800  
 Ile Ile Leu Ala Val Val Phe Ile Arg Ile Lys Lys Asp Lys Gly Lys  
 805 810 815  
 Asp Asn Val Glu Ser Ala Gln Ala Ser Glu Val Lys Pro Leu Arg Ser  
 820 825 830

## (2) INFORMATION FOR SEQ ID NO:179:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1827 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

Met Ala Arg Lys Lys Phe Ser Gly Leu Glu Ile Ser Leu Ile Val Leu  
 1 5 10 15  
 Phe Val Ile Val Thr Ile Ile Ala Ile Ala Leu Ile Val Val Leu Ala  
 20 25 30  
 Thr Lys Thr Pro Ala Val Asp Glu Ile Ser Asp Ser Thr Ser Thr Pro  
 35 40 45  
 Ala Thr Thr Arg Val Thr Thr Asn Pro Ser Asp Ser Gly Lys Cys Pro  
 50 55 60  
 20 Asn Val Leu Asn Asp Pro Val Asn Val Arg Ile Asn Cys Ile Pro Glu  
 65 70 75 80  
 Gln Phe Pro Thr Glu Gly Ile Cys Ala Gln Arg Gly Cys Cys Trp Arg  
 85 90 95  
 Pro Trp Asn Asp Ser Leu Ile Pro Trp Cys Phe Phe Val Asp Asn His  
 100 105 110  
 Gly Tyr Asn Val Gln Asp Met Thr Thr Ser Ile Gly Val Glu Ala  
 115 120 125  
 25 Lys Leu Asn Arg Ile Pro Ser Pro Thr Leu Phe Gly Asn Asp Ile Asn  
 130 135 140  
 Ser Val Leu Phe Thr Thr Gln Asn Gln Thr Pro Asn Arg Phe Arg Phe  
 145 150 155 160  
 Lys Ile Thr Asp Pro Asn Asn Arg Arg Tyr Glu Val Pro His Gln Tyr  
 165 170 175  
 Val Lys Glu Phe Thr Gly Pro Thr Val Ser Asp Thr Leu Tyr Asp Val  
 180 185 190  
 Lys Val Ala Gln Asn Pro Phe Ser Ile Gln Val Ile Arg Lys Ser Asn  
 195 200 205  
 30 Gly Lys Thr Leu Phe Asp Thr Ser Ile Gly Pro Leu Val Tyr Ser Asp  
 210 215 220  
 Gln Tyr Leu Gln Ile Ser Ala Arg Leu Pro Ser Tyr Ile Tyr Gly  
 225 230 235 240  
 Ile Gly Glu Gln Val His Lys Arg Phe Arg His Asp Leu Ser Trp Lys  
 245 250 255  
 Thr Trp Pro Ile Phe Thr Arg Asp Gln Leu Pro Gly Asp Asn Asn Asn  
 260 265 270  
 35 Asn Leu Tyr Gly His Gln Thr Phe Phe Met Cys Ile Glu Asp Thr Ser  
 275 280 285  
 Gly Lys Ser Phe Gly Val Phe Leu Met Asn Ser Asn Ala Met Glu Ile  
 290 295 300  
 Phe Ile Gln Pro Thr Pro Ile Val Thr Tyr Arg Val Thr Gly Gly Ile

305	310	315	320
Leu Asp Phe Tyr Ile Leu Leu Gly Asp Thr Pro Glu Gln Val Val Gln			
325	330	335	
Gln Tyr Gln Gln Leu Val Gly Leu Pro Ala Met Pro Ala Tyr Trp Asn			
340	345	350	
Leu Gly Phe Gln Leu Ser Arg Trp Asn Tyr Lys Ser Leu Asp Val Val			
355	360	365	
5 Lys Glu Val Val Arg Arg Asn Arg Glu Ala Gly Ile Pro Phe Asp Thr			
370	375	380	
Gln Val Thr Asp Ile Asp Tyr Met Glu Asp Lys Lys Asp Phe Thr Tyr			
385	390	395	400
Asp Gln Val Ala Phe Asn Gly Leu Pro Gln Phe Val Gln Asp Leu His			
405	410	415	
Asp His Gly Gln Lys Tyr Val Ile Ile Leu Asp Pro Ala Ile Ser Ile			
420	425	430	
Gly Arg Arg Ala Asn Gly Thr Thr Tyr Ala Thr Tyr Glu Arg Gly Asn			
435	440	445	
10 Thr Gln His Val Trp Ile Asn Glu Ser Asp Gly Ser Thr Pro Ile Ile			
450	455	460	
Gly Glu Val Trp Pro Gly Leu Thr Val Tyr Pro Asp Phe Thr Asn Pro			
465	470	475	480
Asn Cys Ile Asp Trp Trp Ala Asn Glu Cys Ser Ile Phe His Gln Glu			
485	490	495	
Val Gln Tyr Asp Gly Leu Trp Ile Asp Met Asn Glu Val Ser Ser Phe			
500	505	510	
15 Ile Gln Gly Ser Thr Lys Gly Cys Asn Val Asn Lys Leu Asn Tyr Pro			
515	520	525	
Pro Phe Thr Pro Asp Ile Leu Asp Lys Leu Met Tyr Ser Lys Thr Ile			
530	535	540	
Cys Met Asp Ala Val Gln Asn Trp Gly Lys Gln Tyr Asp Val His Ser			
545	550	555	560
Leu Tyr Gly Tyr Ser Met Ala Ile Ala Thr Glu Gln Ala Val Gln Lys			
565	570	575	
Val Phe Pro Asn Lys Arg Ser Phe Ile Leu Thr Arg Ser Thr Phe Ala			
580	585	590	
20 Gly Ser Gly Arg His Ala Ala His Trp Leu Gly Asp Asn Thr Ala Ser			
595	600	605	
Trp Glu Gln Met Glu Trp Ser Ile Thr Gly Met Leu Glu Phe Ser Leu			
610	615	620	
Phe Gly Ile Pro Leu Val Gly Ala Asp Ile Cys Gly Phe Val Ala Glu			
625	630	635	640
Thr Thr Glu Leu Cys Arg Arg Trp Met Gln Leu Gly Ala Phe Tyr			
645	650	655	
25 Pro Phe Ser Arg Asn His Asn Ser Asp Gly Tyr Glu His Gln Asp Pro			
660	665	670	
Ala Phe Phe Gly Gln Asn Ser Leu Leu Val Lys Ser Ser Arg Gln Tyr			
675	680	685	
Leu Thr Ile Arg Tyr Thr Leu Leu Pro Phe Leu Tyr Thr Leu Phe Tyr			
690	695	700	
Lys Ala His Val Phe Gly Glu Thr Val Ala Arg Pro Val Leu His Glu			
705	710	715	720
Phe Tyr Glu Asp Thr Asn Ser Trp Ile Glu Asp Thr Glu Phe Leu Trp			
725	730	735	
30 Gly Pro Ala Leu Leu Ile Thr Pro Val Leu Lys Gln Gly Ala Asp Thr			
740	745	750	
Val Ser Ala Tyr Ile Pro Asp Ala Ile Trp Tyr Asp Tyr Glu Ser Gly			
755	760	765	
Ala Lys Arg Pro Trp Arg Lys Gln Arg Val Asp Met Tyr Leu Pro Ala			
770	775	780	
Asp Lys Ile Gly Leu His Leu Arg Gly Gly Tyr Ile Ile Pro Ile Gln			
785	790	795	800
35 Glu Pro Asp Val Thr Thr Ala Ser Arg Lys Asn Pro Leu Gly Leu			
805	810	815	
Ile Val Ala Leu Gly Glu Asn Asn Thr Ala Lys Gly Asp Phe Phe Trp			
820	825	830	
Asp Asp Gly Glu Thr Lys Asp Thr Ile Gln Asn Gly Asn Tyr Ile Leu			

	835	840	845	
	Tyr Thr Phe Ser Val Ser Asn Asn Thr Leu Asp Ile Val Cys Thr His			
	850	855	860	
	Ser Ser Tyr Gln Glu Gly Thr Thr Leu Ala Phe Gln Thr Val Lys Ile			
	865	870	875	880
	Leu Gly Leu Thr Asp Ser Val Thr Glu Val Arg Val Ala Glu Asn Asn			
	885	890	895	
5	Gln Pro Met Asn Ala His Ser Asn Phe Thr Tyr Asp Ala Ser Asn Gln			
	900	905	910	
	Val Leu Leu Ile Ala Asp Leu Lys Leu Asn Leu Gly Arg Asn Phe Ser			
	915	920	925	
	Val Gln Trp Asn Gln Ile Phe Ser Glu Asn Glu Arg Phe Asn Cys Tyr			
	930	935	940	
	Pro Asp Ala Asp Leu Ala Thr Glu Gln Lys Cys Thr Gln Arg Gly Cys			
	945	950	955	960
	Val Trp Arg Thr Gly Ser Ser Leu Ser Lys Ala Pro Glu Cys Tyr Phe			
	965	970	975	
10	Pro Arg Gln Asp Asn Ser Tyr Ser Val Asn Ser Ala Arg Tyr Ser Ser			
	980	985	990	
	Met Gly Ile Thr Ala Asp Leu Gln Leu Asn Thr Ala Asn Ala Arg Ile			
	995	1000	1005	
	Lys Leu Pro Ser Asp Pro Ile Ser Thr Leu Arg Val Glu Val Lys Tyr			
	1010	1015	1020	
	His Lys Asn Asp Met Leu Gln Phe Lys Ile Tyr Asp Pro Gln Lys Lys			
	025	1030	1035	1040
15	Arg Tyr Glu Val Pro Val Pro Leu Asn Ile Pro Thr Thr Pro Ile Ser			
	1045	1050	1055	
	Thr Tyr Glu Asp Arg Leu Tyr Asp Val Glu Ile Lys Glu Asn Pro Phe			
	1060	1065	1070	
	Gly Ile Gln Ile Arg Arg Ser Ser Gly Arg Val Ile Trp Asp Ser			
	1075	1080	1085	
	Trp Leu Pro Gly Phe Ala Phe Asn Asp Gln Phe Ile Gln Ile Ser Thr			
	1090	1095	1100	
	Arg Leu Pro Ser Glu Tyr Ile Tyr Gly Phe Gly Glu Val Glu His Thr			
20	105	1110	1115	1120
	Ala Phe Lys Arg Asp Leu Asn Trp Asn Thr Trp Gly Met Phe Thr Arg			
	1125	1130	1135	
	Asp Gln Pro Pro Gly Tyr Lys Leu Asn Ser Tyr Gly Phe His Pro Tyr			
	1140	1145	1150	
	Tyr Met Ala Leu Glu Glu Gly Asn Ala His Gly Val Phe Leu Leu			
	1155	1160	1165	
	Asn Ser Asn Ala Met Asp Val Thr Phe Gln Pro Thr Pro Ala Leu Thr			
	1170	1175	1180	
25	Tyr Arg Thr Val Gly Gly Ile Leu Asp Phe Tyr Met Phe Leu Gly Pro			
	1185	1190	1195	1200
	Thr Pro Gln Val Ala Thr Lys Gln Tyr His Glu Val Ile Gly His Pro			
	1205	1210	1215	
	Val Met Pro Ala Tyr Trp Ala Leu Gly Phe Gln Leu Cys Arg Tyr Gly			
	1220	1225	1230	
	Tyr Ala Asn Thr Ser Glu Val Arg Glu Leu Tyr Asp Ala Met Val Ala			
	1235	1240	1245	
	Ala Asn Ile Pro Tyr Asp Val Gln Tyr Thr Asp Ile Asp Tyr Met Glu			
30	1250	1255	1260	
	Arg Gln Leu Asp Phe Thr Ile Gly Glu Ala Phe Gln Asp Leu Pro Gln			
	1265	1270	1275	1280
	Phe Val Asp Lys Ile Arg Gly Glu Gly Met Arg Tyr Ile Ile Leu			
	1285	1290	1295	
	Asp Pro Ala Ile Ser Gly Asn Glu Thr Lys Thr Tyr Pro Ala Phe Glu			
	1300	1305	1310	
	Arg Gly Gln Gln Asn Asp Val Phe Val Lys Trp Pro Asn Thr Asn Asp			
	1315	1320	1325	
35	Ile Cys Trp Ala Lys Val Trp Pro Asp Leu Pro Asn Ile Thr Ile Asp			
	1330	1335	1340	
	Lys Thr Leu Thr Glu Asp Glu Ala Val Asn Ala Ser Arg Ala His Val			
	345	1350	1355	1360
	Ala Phe Pro Asp Phe Arg Thr Ser Thr Ala Glu Trp Trp Ala Arg			

	1365	1370	1375	
	Glu Ile Val Asp Phe Tyr Asn Glu Lys Met Lys Phe Asp Gly Leu Trp			
	1380	1385	1390	
	Ile Asp Met Asn Glu Pro Ser Ser Phe Val Asn Gly Thr Thr Thr Asn			
	1395	1400	1405	
	Gln Cys Arg Asn Asp Glu Leu Asn Tyr Pro Pro Tyr Phe Pro Glu Leu			
	1410	1415	1420	
5	Thr Lys Arg Thr Asp Gly Leu His Phe Arg Thr Ile Cys Met Glu Ala			
	425	1430	1435	1440
	Glu Gln Ile Leu Ser Asp Gly Thr Ser Val Leu His Tyr Asp Val His			
	1445	1450	1455	
	Asn Leu Tyr Gly Trp Ser Gln Met Lys Pro Thr His Asp Ala Leu Gln			
	1460	1465	1470	
	Lys Thr Thr Gly Lys Arg Gly Ile Val Ile Ser Arg Ser Thr Tyr Pro			
	1475	1480	1485	
	Thr Ser Gly Arg Trp Gly Gly His Trp Leu Gly Asp Asn Tyr Ala Arg			
10	1490	1495	1500	
	Trp Asp Asn Met Asp Lys Ser Ile Ile Gly Met Met Glu Phe Ser Leu			
	505	1510	1515	1520
	Phe Gly Ile Ser Tyr Thr Gly Ala Asp Ile Cys Gly Phe Phe Asn Asn			
	1525	1530	1535	
	Ser Glu Tyr His Leu Cys Thr Arg Trp Met Gln Leu Gly Ala Phe Tyr			
	1540	1545	1550	
	Pro Tyr Ser Arg Asn His Asn Ile Ala Asn Thr Arg Arg Gln Asp Pro			
	1555	1560	1565	
15	Ala Ser Trp Asn Glu Thr Phe Ala Glu Met Ser Arg Asn Ile Leu Asn			
	1570	1575	1580	
	Ile Arg Tyr Thr Leu Leu Pro Tyr Phe Tyr Thr Gln Met His Glu Ile			
	585	1590	1595	1600
	His Ala Asn Gly Gly Thr Val Ile Arg Pro Leu Leu His Glu Phe Phe			
	1605	1610	1615	
	Asp Glu Lys Pro Thr Trp Asp Ile Phe Lys Gln Phe Leu Trp Gly Pro			
	1620	1625	1630	
	Ala Phe Met Val Thr Pro Val Leu Glu Pro Tyr Val Gln Thr Val Asn			
20	1635	1640	1645	
	Ala Tyr Val Pro Asn Ala Arg Trp Phe Asp Tyr His Thr Gly Lys Asp			
	1650	1655	1660	
	Ile Gly Val Arg Gly Gln Phe Gln Thr Phe Asn Ala Ser Tyr Asp Thr			
	665	1670	1675	1680
	Ile Asn Leu His Val Arg Gly Gly His Ile Leu Pro Cys Gln Glu Pro			
	1685	1690	1695	
	Ala Gln Asn Thr Phe Tyr Ser Arg Gln Lys His Met Lys Leu Ile Val			
	1700	1705	1710	
25	Ala Ala Asp Asp Asn Gln Met Ala Gln Gly Ser Leu Phe Trp Asp Asp			
	1715	1720	1725	
	Gly Glu Ser Ile Asp Thr Tyr Glu Arg Asp Leu Tyr Leu Ser Val Gln			
	1730	1735	1740	
	Phe Asn Leu Asn Gln Thr Thr Leu Thr Ser Thr Ile Leu Lys Arg Gly			
	745	1750	1755	1760
	Tyr Ile Asn Lys Ser Glu Thr Arg Leu Gly Ser Leu His Val Trp Gly			
	1765	1770	1775	
	Lys Gly Thr Thr Pro Val Asn Ala Val Thr Leu Thr Tyr Asn Gly Asn			
30	1780	1785	1790	
	Lys Asn Ser Leu Pro Phe Asn Glu Asp Thr Thr Asn Met Ile Leu Arg			
	1795	1800	1805	
	Ile Asp Leu Thr Thr His Asn Val Thr Leu Glu Glu Pro Ile Glu Ile			
	1810	1815	1820	
	Asn Trp Ser			
	825			

(2) INFORMATION FOR SEQ ID NO:180:

35

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2284 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA  
(ix) FEATURE:(A) NAME/KEY: Coding Sequence  
(B) LOCATION: 45...2099  
5 (D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

	GCCTTACTGC AGGAAGGCAC TCCGAAGACA TAAGTCGGTG AGAC ATG GCT GAA GAT	56
	Met Ala Glu Asp	
	1	
10	AAA AGC AAG AGA GAC TCC ATC GAG ATG AGT ATG AAG GGA TGC CAG ACA	104
	Lys Ser Lys Arg Asp Ser Ile Glu Met Ser Met Lys Gly Cys Gln Thr	
	5 10 15 20	
	AAC AAC GGG TTT GTC CAT AAT GAA GAC ATT CTG GAG CAG ACC CCG GAT	152
	Asn Asn Gly Phe Val His Asn Glu Asp Ile Leu Glu Gln Thr Pro Asp	
	25 30 35	
15	CCA GGC AGC TCA ACA GAC AAC CTG AAG CAC AGC ACC AGG GGC ATC CTT	200
	Pro Gly Ser Ser Thr Asp Asn Leu Lys His Ser Thr Arg Gly Ile Leu	
	40 45 50	
	GGC TCC CAG GAG CCC GAC TTC AAG GGC GTC CAG CCC TAT GCG GGG ATG	248
	Gly Ser Gln Glu Pro Asp Phe Lys Gly Val Gln Pro Tyr Ala Gly Met	
	55 60 65	
20	CCC AAG GAG GTG CTG TTC CAG TCT GGC CAG GCC CGC TAC CGC ATA	296
	Pro Lys Glu Val Leu Phe Gln Phe Ser Gly Gln Ala Arg Tyr Arg Ile	
	70 75 80	
	CCT CGG GAG ATC CTC TTC TGG CTC ACA GTG GCT TCT GTG CTG GTG CTC	344
	Pro Arg Glu Ile Leu Phe Trp Leu Thr Val Ala Ser Val Leu Val Leu	
	85 90 95 100	
	ATC GCG GCC ACC ATA GCC ATC ATT GCC CTC TCT CCA AAG TGC CTA GAC	392
	Ile Ala Ala Thr Ile Ala Ile Ala Leu Ser Pro Lys Cys Leu Asp	
	105 110 115	
25	TGG TGG CAG GAG GGG CCC ATG TAC CAG ATC TAC CCA AGG TCT TTC AAG	440
	Trp Trp Gln Glu Gly Pro Met Tyr Gln Ile Tyr Pro Arg Ser Phe Lys	
	120 125 130	
	GAC AGT AAC AAG GAT GGG AAC GGA GAT CTG AAA GGT ATT CAA GAT AAA	488
	Asp Ser Asn Lys Asp Gly Asn Gly Asp Leu Lys Gly Ile Gln Asp Lys	
	135 140 145	
30	CTG GAC TAC ATC ACA GCT TTA AAT ATA AAA ACT GTT TGG ATT ACT TCA	536
	Leu Asp Tyr Ile Thr Ala Leu Asn Ile Lys Thr Val Trp Ile Thr Ser	
	150 155 160	
	TTT TAT AAA TCG TCC CTT AAA GAT TTC AGA TAT GGT GTT GAA GAT TTC	584
	Phe Tyr Lys Ser Ser Leu Lys Asp Phe Arg Tyr Gly Val Glu Asp Phe	
	165 170 175 180	
35	CGG GAA GTT GAT CCC ATT TTT GGA ACG ATG GAA GAT TTT GAG AAT CTG	632
	Arg Glu Val Asp Pro Ile Phe Gly Thr Met Glu Asp Phe Glu Asn Leu	
	185 190 195	
	GTT GCA GCC ATA CAT GAT AAA GGT TTA AAA TTA ATC ATC GAT TTC ATA	680
	Val Ala Ala Ile His Asp Lys Gly Leu Lys Leu Ile Ile Asp Phe Ile	

	200	205	210	
	CCA AAC CAC ACG AGT GAT AAA CAT ATT TGG TTT CAA TTG AGT CGG ACA			728
	Pro Asn His Thr Ser Asp Lys His Ile Trp Phe Gln Leu Ser Arg Thr			
	215	220	225	
5	CGG ACA GGA AAA TAT ACT GAT TAT TAT ATC TGG CAT GAC TGT ACC CAT			776
	Arg Thr Gly Lys Tyr Thr Asp Tyr Tyr Ile Trp His Asp Cys Thr His			
	230	235	240	
	GAA AAT GGC AAA ACC ATT CCA CCC AAC AAC TGG TTA AGT GTG TAT GGA			824
	Glu Asn Gly Lys Thr Ile Pro Pro Asn Asn Trp Leu Ser Val Tyr Gly			
	245	250	255	260
10	AAC TCC AGT TGG CAC TTT GAC GAA GTG CGA AAC CAA TGT TAT TTT CAT			872
	Asn Ser Ser Trp His Phe Asp Glu Val Arg Asn Gln Cys Tyr Phe His			
	265	270	275	
	CAG TTT ATG AAA GAG CAA CCT GAT TTA AAT TTC CGC AAT CCT GAT GTT			920
	Gln Phe Met Lys Glu Gln Pro Asp Leu Asn Phe Arg Asn Pro Asp Val			
	280	285	290	
	CAA GAA GAA ATA AAA GAA ATT TTA CGG TTC TGG CTC ACA AAG GGT GTT			968
	Gln Glu Glu Ile Lys Glu Ile Leu Arg Phe Trp Leu Thr Lys Gly Val			
	295	300	305	
15	GAT GGT TTT AGT TTG GAT GCT GTT AAA TTC CTC CTA GAA GCA AAG CAC			1016
	Asp Gly Phe Ser Leu Asp Ala Val Lys Phe Leu Leu Glu Ala Lys His			
	310	315	320	
	CTG AGA GAT GAG ATC CAA GTA AAT AAG ACC CAA ATC CCG GAC ACG GTC			1064
	Leu Arg Asp Glu Ile Gln Val Asn Lys Thr Gln Ile Pro Asp Thr Val			
	325	330	335	340
20	ACA CAA TAC TCG GAG CTG TAC CAT GAC TTC ACC ACC ACG CAG GTG GGA			1112
	Thr Gln Tyr Ser Glu Leu Tyr His Asp Phe Thr Thr Thr Gln Val Gly			
	345	350	355	
	ATG CAC GAC ATT GTC CGC AGC TTC CCG CAG ACC ATG GAC CAA TAC AGC			1160
	Met His Asp Ile Val Arg Ser Phe Arg Gln Thr Met Asp Gln Tyr Ser			
	360	365	370	
25	ACG GAG CCC GGC AGA TAC AGG TTC ATG GGG ACT GAA GCC TAT GCA GAG			1208
	Thr Glu Pro Gly Arg Tyr Arg Phe Met Gly Thr Glu Ala Tyr Ala Glu			
	375	380	385	
	AGT ATT GAC AGG ACC GTG ATG TAC TAT GGA TTG CCA TTT ATC CAA GAA			1256
	Ser Ile Asp Arg Thr Val Met Tyr Tyr Gly Leu Pro Phe Ile Gln Glu			
	390	395	400	
30	GCT GAT TTT CCC TTC AAC AAT TAC CTC AGC ATG CTA GAC ACT GTT TCT			1304
	Ala Asp Phe Pro Phe Asn Asn Tyr Leu Ser Met Leu Asp Thr Val Ser			
	405	410	415	420
	GGG AAC AGC GTG TAT GAG GTT ATC ACA TCC TGG ATG GAA AAC ATG CCA			1352
	Gly Asn Ser Val Tyr Glu Val Ile Thr Ser Trp Met Glu Asn Met Pro			
	425	430	435	
	GAA GGA AAA TGG CCT AAC TGG ATT GGT GGA CCA GAC AGT TCA CGG			1400
	Glu Gly Lys Trp Pro Asn Trp Met Ile Gly Gly Pro Asp Ser Ser Arg			
	440	445	450	
35	CTG ACT TCG CGT TTG GGG AAT CAG TAT GTC AAC GTG ATG AAC ATG CTT			1448
	Leu Thr Ser Arg Leu Gly Asn Gln Tyr Val Asn Val Met Asn Met Leu			
	455	460	465	

CTT	TTC	ACA	CTC	CCT	GGA	ACT	CCT	ATA	ACT	TAC	TAT	GGA	GAA	GAA	ATT	1496	
Leu	Phe	Thr	Leu	Pro	Gly	Thr	Pro	Ile	Thr	Tyr	Tyr	Gly	Glu	Glu	Ile		
470												480					
GGA	ATG	GGA	AAT	ATT	GTA	GCC	GCA	AAT	CTC	AAT	GAA	AGC	TAT	GAT	ATT	1544	
Gly	Met	Gly	Asn	Ile	Val	Ala	Ala	Asn	Leu	Asn	Glu	Ser	Tyr	Asp	Ile		
485											495				500		
5	AAT	ACC	CTT	CGC	TCA	AAG	TCA	CCA	ATG	CAG	TGG	GAC	AAT	AGT	TCA	AAT	1592
Asn	Thr	Leu	Arg	Ser	Ser	Pro	Met	Gln	Trp	Asp	Asn	Ser	Ser	Asn			
											510			515			
GCT	GGT	TTT	TCT	GAA	GCT	AGT	AAC	ACC	TGG	TTA	CCT	ACC	AAT	TCA	GAT	1640	
Ala	Gly	Phe	Ser	Glu	Ala	Ser	Asn	Thr	Trp	Leu	Pro	Thr	Asn	Ser	Asp		
											520	525		530			
10	TAC	CAC	ACT	GTG	AAT	GTT	GAT	GTC	CAA	AAG	ACT	CAG	CCC	AGA	TCG	GCT	1688
Tyr	His	Thr	Val	Asn	Val	Asp	Val	Gln	Lys	Thr	Gln	Pro	Arg	Ser	Ala		
											535	540		545			
TTG	AAG	TTA	TAT	CAA	GAT	TTA	AGT	CTA	CTT	CAT	GCC	AAT	GAG	CTA	CTC	1736	
Leu	Lys	Leu	Tyr	Gln	Asp	Leu	Ser	Leu	Leu	His	Ala	Asn	Glu	Leu	Leu		
											550	555		560			
15	CTC	AAC	AGG	GGC	TGG	TTT	TGC	CAT	TTG	AGG	AAT	GAC	AGC	CAC	TAT	GTT	1784
Leu	Asn	Arg	Gly	Trp	Phe	Cys	His	Leu	Arg	Asn	Asp	Ser	His	Tyr	Val		
											565	570		575		580	
GTG	TAC	ACA	AGA	GAG	CTG	GAT	GGC	ATC	GAC	AGA	ATC	TTT	ATC	GTG	GTT	1832	
Val	Tyr	Thr	Arg	Glu	Leu	Asp	Gly	Ile	Asp	Arg	Ile	Phe	Ile	Val	Val		
											585	590		595			
20	CTG	AAT	TTT	GGA	GAA	TCA	ACA	CTG	TTA	AAT	CTA	CAT	AAT	ATG	ATT	TCG	1880
Leu	Asn	Phe	Gly	Glu	Ser	Thr	Leu	Leu	Asn	Leu	His	Asn	Met	Ile	Ser		
											600	605		610			
GGC	CTT	CCC	GCT	AAA	ATA	AGA	ATA	AGG	TTA	AGT	ACC	AAT	TCT	GCC	GAC	1928	
Gly	Leu	Pro	Ala	Lys	Ile	Arg	Ile	Arg	Leu	Ser	Thr	Asn	Ser	Ala	Asp		
											615	620		625			
25	AAA	GGC	AGT	AAA	GTT	GAT	ACA	AGT	GGC	ATT	TTT	CTG	GAC	AAG	GGA	GAG	1976
Lys	Gly	Ser	Lys	Val	Asp	Thr	Ser	Gly	Ile	Phe	Leu	Asp	Lys	Gly	Glu		
											630	635		640			
GGA	CTC	ATC	TTT	GAA	CAC	AAC	ACG	AAG	AAT	CTC	CTT	CAT	CGC	CAA	ACA	2024	
Gly	Leu	Ile	Phe	Glu	His	Asn	Thr	Lys	Asn	Leu	Leu	His	Arg	Gln	Thr		
											645	650		655		660	
GCT	TTC	AGA	GAT	AGA	TGC	TTT	GTT	TCC	AAT	CGA	GCA	TGC	TAT	TCC	AGT	2072	
Ala	Phe	Arg	Asp	Arg	Cys	Phe	Val	Ser	Asn	Arg	Ala	Cys	Tyr	Ser	Ser		
											665	670		675			
30	GTA	CTG	AAC	ATA	CTG	TAT	ACC	TCG	TGT	TAGGCACCTT	TATGAAGAGA	TGAAGAC				2126	
Val	Leu	Asn	Ile	Leu	Tyr	Thr	Ser	Cys									
											680	685					
ACTGGCATTT	CAGTGGGATT	GTAAGCATT	T	GTAATAGCTT	CATGTACAGC	ATGCTGCTTG										2186	
GTGAACAAATC	ATTAATTCTT	CGATATTCT	GTAGCTTGAA	TGTAACCGCT	TTAAGAAAGG											2246	
TTCTCAAATG	TTTTGAAAAAA	AATAAAATGT	TTAAAGAGT													2284	

(2) INFORMATION FOR SEQ ID NO:181:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 685 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

	Met	Ala	Glu	Asp	Lys	Ser	Lys	Arg	Asp	Ser	Ile	Glu	Met	Ser	Met	Lys
5	1			5				10				15				
	Gly	Cys	Gln	Thr	Asn	Asn	Gly	Phe	Val	His	Asn	Glu	Asp	Ile	Leu	Glu
				20				25				30				
	Gln	Thr	Pro	Asp	Pro	Gly	Ser	Ser	Thr	Asp	Asn	Leu	Lys	His	Ser	Thr
				35				40				45				
	Arg	Gly	Ile	Leu	Gly	Ser	Gln	Glu	Pro	Asp	Phe	Lys	Gly	Val	Gln	Pro
				50				55				60				
	Tyr	Ala	Gly	Met	Pro	Lys	Glu	Val	Leu	Phe	Gln	Phe	Ser	Gly	Gln	Ala
				65				70				75				80
10	Arg	Tyr	Arg	Ile	Pro	Arg	Glu	Ile	Leu	Phe	Trp	Leu	Thr	Val	Ala	Ser
								85				90				95
	Val	Leu	Val	Leu	Ile	Ala	Ala	Thr	Ile	Ala	Ile	Ile	Ala	Leu	Ser	Pro
								100				105				110
	Lys	Cys	Leu	Asp	Trp	Trp	Gln	Glu	Gly	Pro	Met	Tyr	Gln	Ile	Tyr	Pro
								115				120				125
	Arg	Ser	Phe	Lys	Asp	Ser	Asn	Lys	Asp	Gly	Asn	Gly	Asp	Leu	Lys	Gly
								130				135				140
	Ile	Gln	Asp	Lys	Leu	Asp	Tyr	Ile	Thr	Ala	Leu	Asn	Ile	Lys	Thr	Val
15								145				150				160
	Trp	Ile	Thr	Ser	Phe	Tyr	Lys	Ser	Ser	Leu	Lys	Asp	Phe	Arg	Tyr	Gly
												165				175
	Val	Glu	Asp	Phe	Arg	Glu	Val	Asp	Pro	Ile	Phe	Gly	Thr	Met	Glu	Asp
								180				185				190
	Phe	Glu	Asn	Leu	Val	Ala	Ala	Ile	His	Asp	Lys	Gly	Leu	Lys	Leu	Ile
								195				200				205
	Ile	Asp	Phe	Ile	Pro	Asn	His	Thr	Ser	Asp	Lys	His	Ile	Trp	Phe	Gln
								210				215				220
20	Leu	Ser	Arg	Thr	Arg	Thr	Gly	Lys	Tyr	Thr	Asp	Tyr	Tyr	Ile	Trp	His
								225				230				240
	Asp	Cys	Thr	His	Glu	Asn	Gly	Lys	Thr	Ile	Pro	Pro	Asn	Asn	Trp	Leu
								245				250				255
	Ser	Val	Tyr	Gly	Asn	Ser	Ser	Trp	His	Phe	Asp	Glu	Val	Arg	Asn	Gln
								260				265				270
	Cys	Tyr	Phe	His	Gln	Phe	Met	Lys	Glu	Gln	Pro	Asp	Leu	Asn	Phe	Arg
								275				280				285
	Asn	Pro	Asp	Val	Gln	Glu	Glu	Ile	Lys	Glu	Ile	Leu	Arg	Phe	Trp	Leu
25								290				295				300
	Thr	Lys	Gly	Val	Asp	Gly	Phe	Ser	Leu	Asp	Ala	Val	Lys	Phe	Leu	Leu
								305				310				320
	Glu	Ala	Lys	His	Leu	Arg	Asp	Glu	Ile	Gln	Val	Asn	Lys	Thr	Gln	Ile
								325				330				335
	Pro	Asp	Thr	Val	Thr	Gln	Tyr	Ser	Glu	Leu	Tyr	His	Asp	Phe	Thr	Thr
								340				345				350
	Thr	Gln	Val	Gly	Met	His	Asp	Ile	Val	Arg	Ser	Phe	Arg	Gln	Thr	Met
								355				360				365
30	Asp	Gln	Tyr	Ser	Thr	Glu	Pro	Gly	Arg	Tyr	Arg	Phe	Met	Gly	Thr	Glu
								370				375				380
	Ala	Tyr	Ala	Glu	Ser	Ile	Asp	Arg	Thr	Val	Met	Tyr	Tyr	Gly	Leu	Pro
								385				390				400
	Phe	Ile	Gln	Glu	Ala	Asp	Phe	Pro	Phe	Asn	Asn	Tyr	Leu	Ser	Met	Leu
								405				410				415
	Asp	Thr	Val	Ser	Gly	Asn	Ser	Val	Tyr	Glu	Val	Ile	Thr	Ser	Trp	Met
								420				425				430
	Glu	Asn	Met	Pro	Glu	Gly	Lys	Trp	Pro	Asn	Trp	Met	Ile	Gly	Gly	Pro
35								435				440				445
	Asp	Ser	Ser	Arg	Leu	Thr	Ser	Arg	Leu	Gly	Asn	Gln	Tyr	Val	Asn	Val
								450				455				460
	Met	Asn	Met	Leu	Leu	Phe	Thr	Leu	Pro	Gly	Thr	Pro	Ile	Thr	Tyr	Tyr
								465				470				480

Gly Glu Glu Ile Gly Met Gly Asn Ile Val Ala Ala Asn Leu Asn Glu  
 485 490 495  
 Ser Tyr Asp Ile Asn Thr Leu Arg Ser Lys Ser Pro Met Gln Trp Asp  
 500 505 510  
 Asn Ser Ser Asn Ala Gly Phe Ser Glu Ala Ser Asn Thr Trp Leu Pro  
 515 520 525  
 Thr Asn Ser Asp Tyr His Thr Val Asn Val Asp Val Gln Lys Thr Gln  
 530 535 540  
 5 Pro Arg Ser Ala Leu Lys Leu Tyr Gln Asp Leu Ser Leu Leu His Ala  
 545 550 555 560  
 Asn Glu Leu Leu Asn Arg Gly Trp Phe Cys His Leu Arg Asn Asp  
 565 570 575  
 Ser His Tyr Val Val Tyr Thr Arg Glu Leu Asp Gly Ile Asp Arg Ile  
 580 585 590  
 Phe Ile Val Val Leu Asn Phe Gly Glu Ser Thr Leu Leu Asn Leu His  
 595 600 605  
 10 Asn Met Ile Ser Gly Leu Pro Ala Lys Ile Arg Ile Arg Leu Ser Thr  
 610 615 620  
 Asn Ser Ala Asp Lys Gly Ser Lys Val Asp Thr Ser Gly Ile Phe Leu  
 625 630 635 640  
 Asp Lys Gly Glu Gly Leu Ile Phe Glu His Asn Thr Lys Asn Leu Leu  
 645 650 655  
 His Arg Gln Thr Ala Phe Arg Asp Arg Cys Phe Val Ser Asn Arg Ala  
 660 665 670  
 Cys Tyr Ser Ser Val Leu Asn Ile Leu Tyr Thr Ser Cys  
 15 675 680 685

## (2) INFORMATION FOR SEQ ID NO:182:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 54 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

Leu Val Pro Arg Gly Ser Pro Gly Ile Pro Gly Ser Arg Val Gly Gln  
 1 5 10 15  
 Cys Thr Asp Ser Asp Val Arg Arg Pro Trp Ala Arg Ser Cys Ala His  
 20 25 30  
 25 Gln Gly Cys Gly Ala Gly Thr Arg Asn Ser His Gly Cys Ile Thr Arg  
 35 40 45  
 Pro Leu Arg Gln Ala Ser  
 50

## (2) INFORMATION FOR SEQ ID NO:183:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

Ser Ala Arg Asp Ser Gly Pro Ala Glu Asp Gly Ser Arg Ala Val Arg  
 35 1 5 10 15  
 Leu Asn Gly

## (2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

Asp Gly Ser Arg Ala Val Arg Leu Asn Gly Val Glu Asn Ala Asn Thr  
1 5 10 15  
Arg Lys Ser Ser Arg  
20

(2) INFORMATION FOR SEQ ID NO:185:

10

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

Glu Asn Ala Asn Thr Arg Lys Ser Ser Arg Ser Asn Pro Arg Gly Arg  
1 5 10 15  
Arg His Pro

(2) INFORMATION FOR SEQ ID NO:186:

20

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

25

Thr Arg Lys Ser Ser Arg Ser Asn Pro Arg Gly  
1 5 10

(2) INFORMATION FOR SEQ ID NO:187:

30

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

35

Ser Arg Pro Tyr Ser Val Asp Ser Asp Ser Asp Thr Asn Ala Lys His  
1 5 10 15  
Ser Ser His Asn Arg  
20

(2) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

Thr Asn Ala Lys His Ser Ser His Asn Arg Arg Arg Leu Arg Thr Arg Ser  
1 5 10 15  
Arg Pro Asn

(2) INFORMATION FOR SEQ ID NO:189:

10

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

Arg Tyr Lys His Asp Ile Gly Cys Asp Ala Gly Val Asp Lys Lys Ser  
1 5 10 15  
Ser Ser Val Arg Gly Gly Cys Gly  
20

(2) INFORMATION FOR SEQ ID NO:190:

20

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

25

Gly Cys Asp Ala Gly Val Asp Lys Lys Ser Ser Ser Val Arg Gly Gly  
1 5 10 15  
Cys Gly Ala His Ser Ser Pro Pro Arg Ala  
20 25

(2) INFORMATION FOR SEQ ID NO:191:

30

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

35

Gly Ala His Ser Ser Pro Pro Arg Ala Gly Arg Gly Pro Arg Gly Thr  
1 5 10 15  
Met Val Ser Arg Leu  
20

## (2) INFORMATION FOR SEQ ID NO:192:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

5

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

Glu Asn Ala Asn Thr Arg Lys Ser Ser Arg  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:193:

10

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

Lys Lys Arg Ile Ala Gly Leu Pro Trp Tyr Arg Cys Arg Thr Val Ala  
1 5 10 15  
Phe Glu Thr Gly Met Gln Asn Thr Gln Leu Cys Ser Thr Ile Val Gln  
20 25 30  
Leu Ser Phe Thr Pro Glu Glu  
35

20

## (2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

Arg Lys Ser Ser Arg Ser Asn Pro Arg Gly  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:195:

30

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

35

Ser Asn Pro Arg Gly Arg Arg His Pro  
1 5

## (2) INFORMATION FOR SEQ ID NO:196:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

Thr Asn Ala Lys His Ser Ser His Asn  
1 5

## (2) INFORMATION FOR SEQ ID NO:197:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

15 Ser Ser His Asn Arg Arg Leu Arg Thr Arg  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:198:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

20

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

Arg Arg Leu Arg Thr Arg Ser Arg Pro Asn  
1 5 10

25

## (2) INFORMATION FOR SEQ ID NO:199:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

Arg Val Gly Gln Cys Thr Asp Ser Asp Val Arg Arg Pro Trp Ala Arg  
1 5 10 15  
Ser Cys Ala

35

## (2) INFORMATION FOR SEQ ID NO:200:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

5 Val Arg Arg Pro Trp Ala Arg Ser Cys Ala His Gln Gly Cys Gly Ala  
 1 5 10 15  
 Gly Thr Arg Asn Ser  
 20

(2) INFORMATION FOR SEQ ID NO:201:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

10

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

Gly Thr Arg Asn Ser His Gly Cys Ile Thr Arg Pro Leu Arg Gln Ala  
 1 5 10 15  
 15 Ser Gln His

(2) INFORMATION FOR SEQ ID NO:202:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

20

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

Ser Thr Pro Pro Ser Arg Glu Ala Tyr Ser Arg Pro Tyr Ser Val Asp  
 1 5 10 15  
 Ser Asp Ser Asp Thr Met Ala Lys His Ser Ser His Asn Arg Arg Leu  
 25 20 25 30  
 Arg Thr Arg Ser Arg Pro Asn Gly  
 35 40

(2) INFORMATION FOR SEQ ID NO:203:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

30

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

Tyr Ser Lys Val  
 1

35

(2) INFORMATION FOR SEQ ID NO:204:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

5 Phe Pro His Leu  
1

(2) INFORMATION FOR SEQ ID NO:205:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

Tyr Arg Gly Val  
1

15 (2) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

Tyr Gln Thr Ile  
1

(2) INFORMATION FOR SEQ ID NO:207:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

30 Thr Glu Gln Phe  
1

(2) INFORMATION FOR SEQ ID NO:208:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

Thr Glu Val Met  
1

## (2) INFORMATION FOR SEQ ID NO:209:

5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

10 Thr Ser Ala Phe  
1

## (2) INFORMATION FOR SEQ ID NO:210:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

Tyr Thr Arg Phe  
1

## 20 (2) INFORMATION FOR SEQ ID NO:211:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 717 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA  
 (ix) FEATURE:

(A) NAME/KEY: Coding Sequence  
 (B) LOCATION: 1...714  
 (D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

30	ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro	48
	1 5 10 15	
35	ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG Thr Arg Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu	96
	20 25 30	
40	TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu	144
	35 40 45	
45	GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA	192

	Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys	
	50 55 60	
	TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC	240
	Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn	
	65 70 75 80	
5	ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA	288
	Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu	
	85 90 95	
	GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT	336
	Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser	
	100 105 110	
10	AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA	384
	Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu	
	115 120 125	
	ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT	432
	Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn	
	130 135 140	
15	GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT	480
	Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp	
	145 150 155 160	
	GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA	528
	Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu	
	165 170 175	
	GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC	576
	Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr	
	180 185 190	
20	TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC	624
	Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala	
	195 200 205	
	ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG GTT CCG CGT	672
	Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg	
	210 215 220	
25	GGA TCC CCA GGA ATT CCC GGG TCG ACT CGA GCG GCC GCA TCG TGA	717
	Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser	
	225 230 235	

## (2) INFORMATION FOR SEQ ID NO:212:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 238 amino acids

30 (B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

	Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro	
	1 5 10 15	
35	Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu	
	20 25 30	
	Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu	
	35 40 45	

Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 5 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 10 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser  
 225 230 235

15 (2) INFORMATION FOR SEQ ID NO:213:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 25 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 30 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 35 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ser Ser Gln

225	230	235	240
Gly Ser Lys Gln Cys Met Gln Tyr Arg Thr	Gly Arg Leu Thr Val	Gly	
245	250	255	
Ser Glu Tyr Gly Cys Gly Met Asn Pro Ala Arg His Ala	Thr Pro Ala		
260	265	270	
Tyr Pro Ala Arg Leu Leu Pro Arg Tyr Arg			
275	280		

5

(2) INFORMATION FOR SEQ ID NO:214:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 282 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

10

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro			
1	5	10	15
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu			
20	25	30	
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu			
35	40	45	
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys			
50	55	60	
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn			
65	70	75	80
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
85	90	95	
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
100	105	110	
20 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
115	120	125	
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn			
130	135	140	
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
145	150	155	160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165	170	175	
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180	185	190	
25 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195	200	205	
Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
210	215	220	
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Asp			
225	230	235	240
His Ala Leu Gly Thr Asn Leu Arg Ser Asp Asn Ala Lys Glu Pro Gly			
245	250	255	
30 Asp Tyr Asn Cys Cys Gly Asn Gly Asn Ser Thr Gly Arg Lys Val Phe			
260	265	270	
Asn Arg Arg Arg Pro Ser Ala Ile Pro Thr			
275	280		

(2) INFORMATION FOR SEQ ID NO:215:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 279 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 5 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 10 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 15 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Pro  
 225 230 235 240  
 Cys Gly Gly Ser Trp Gly Arg Phe Met Gln Gly Gly Leu Phe Gly Gly  
 245 250 255  
 20 Arg Thr Asp Gly Cys Gly Ala His Arg Asn Arg Thr Ser Ala Ser Leu  
 260 265 270  
 Glu Pro Pro Ser Ser Asp Tyr  
 275

## (2) INFORMATION FOR SEQ ID NO:216:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 277 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 30 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 35 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu

	115	120	125	
	Met Leu Lys Met Phe Glu Asp Arg	Leu Cys His Lys Thr Tyr	Leu Asn	
	130	135	140	
	Gly Asp His Val Thr His Pro Asp Phe Met	Leu Tyr Asp Ala Leu Asp		
	145	150	155	160
	Val Val Leu Tyr Met Asp Pro Met Cys	Leu Asp Ala Phe Pro Lys Leu		
	165	170	175	
5	Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
	180	185	190	
	Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
	195	200	205	
	Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
	210	215	220	
	Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ser Arg Gly			
	225	230	235	240
	Ser Thr Gly Thr Ala Gly Gly Glu Arg Ser Gly Val Leu Asn Leu His			
10	245	250	255	
	Thr Arg Asp Asn Ala Ser Gly Ser Gly Phe Lys Pro Trp Tyr Pro Ser			
	260	265	270	
	Asn Arg Gly His Lys			
	275			

## (2) INFORMATION FOR SEQ ID NO:217:

## (i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 277 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

20	Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro			
	1	5	10	15
	Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu			
	20	25	30	
	Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu			
	35	40	45	
	Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys			
	50	55	60	
	Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn			
25	65	70	75	80
	Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
	85	90	95	
	Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
	100	105	110	
	Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
	115	120	125	
	Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn			
	130	135	140	
30	Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
	145	150	155	160
	Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
	165	170	175	
	Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
	180	185	190	
	Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
	195	200	205	
	Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
35	210	215	220	
	Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ser Ser His			
	225	230	235	240
	Ser Gly Gly Met Asn Arg Ala Tyr Gly Asp Val Phe Arg Glu Leu Arg			
	245	250	255	

Asp Arg Trp Asn Ala Thr Ser His His Thr Arg Pro Thr Pro Gln Leu  
 260 265 270  
 Pro Arg Gly Pro Asn  
 275

## (2) INFORMATION FOR SEQ ID NO:218:

5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 248 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

10 Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 15 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 20 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 25 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser His  
 225 230 235 240  
 Ser Gly Gly Met Asn Arg Ala Tyr  
 245

## (2) INFORMATION FOR SEQ ID NO:219:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 248 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

35 Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30

Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 5 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 10 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Gly Asp  
 225 230 235 240  
 15 Val Phe Arg Glu Leu Arg Asp Arg  
 245

## (2) INFORMATION FOR SEQ ID NO:220:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 248 amino acids  
 (B) TYPE: amino acid  
 20 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 25 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 30 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 35 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala

195	200	205
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg		
210	215	220
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Trp Asn		
225	230	235
Ala Thr Ser His His Thr Arg Pro		
	245	

5

(2) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 247 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

10

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro		
1	5	10
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu		
20	25	30
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu		
35	40	45
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys		
50	55	60
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn		
65	70	75
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu		
85	90	95
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser		
100	105	110
20 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu		
115	120	125
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn		
130	135	140
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp		
145	150	155
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu		
165	170	175
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr		
180	185	190
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala		
195	200	205
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg		
210	215	220
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Thr Pro		
225	230	235
Gln Leu Pro Arg Gly Pro Asn		
	245	

30

(2) INFORMATION FOR SEQ ID NO:222:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 258 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro

1	5	10	15													
Thr	Arg	Leu	Leu	Leu	Glu	Tyr	Leu	Glu	Glu	Lys	Tyr	Glu	Glu	His	Leu	
		20			25					30						
Tyr	Glu	Arg	Asp	Glu	Gly	Asp	Lys	Trp	Arg	Asn	Lys	Lys	Phe	Glu	Leu	
		35			40					45						
Gly	Leu	Glu	Phe	Pro	Asn	Leu	Pro	Tyr	Tyr	Ile	Asp	Gly	Asp	Val	Lys	
		50			55					60						
5	Leu	Thr	Gln	Ser	Met	Ala	Ile	Ile	Arg	Tyr	Ile	Ala	Asp	Lys	His	Asn
		65			70					75				80		
Met	Leu	Gly	Gly	Cys	Pro	Lys	Glu	Arg	Ala	Glu	Ile	Ser	Met	Leu	Glu	
		85								90				95		
Gly	Ala	Val	Leu	Asp	Ile	Arg	Tyr	Gly	Val	Ser	Arg	Ile	Ala	Tyr	Ser	
		100							105				110			
Lys	Asp	Phe	Glu	Thr	Leu	Lys	Val	Asp	Phe	Leu	Ser	Lys	Leu	Pro	Glu	
		115							120				125			
Met	Leu	Lys	Met	Phe	Glu	Asp	Arg	Leu	Cys	His	Lys	Thr	Tyr	Leu	Asn	
		130			135					140						
10	Gly	Asp	His	Val	Thr	His	Pro	Asp	Phe	Met	Leu	Tyr	Asp	Ala	Leu	Asp
		145				150				155					160	
Val	Val	Leu	Tyr	Met	Asp	Pro	Met	Cys	Leu	Asp	Ala	Phe	Pro	Lys	Leu	
		165							170				175			
Val	Cys	Phe	Lys	Arg	Ile	Glu	Ala	Ile	Pro	Gln	Ile	Asp	Lys	Tyr		
		180				185				190						
Leu	Lys	Ser	Ser	Lys	Tyr	Ile	Ala	Trp	Pro	Leu	Gln	Gly	Trp	Gln	Ala	
		195				200					205					
15	Thr	Phe	Gly	Gly	Asp	His	Pro	Pro	Lys	Ser	Asp	Leu	Val	Pro	Arg	
		210			215					220						
Gly	Ser	Pro	Gly	Ile	Pro	Gly	Ser	Thr	Arg	Ala	Ala	Ala	Ser	Gly	Asp	
		225			230					235				240		
Val	Phe	Arg	Glu	Leu	Arg	Asp	Arg	Trp	Asn	Ala	Thr	Ser	His	His	Thr	
		245							250				255			
	Arg	Pro														

## 20 (2) INFORMATION FOR SEQ ID NO:223:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 257 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

Met	Ser	Pro	Ile	Leu	Gly	Tyr	Trp	Lys	Ile	Lys	Gly	Leu	Val	Gln	Pro	
1			5		10					15						
Thr	Arg	Leu	Leu	Leu	Glu	Tyr	Leu	Glu	Glu	Lys	Tyr	Glu	Glu	His	Leu	
		20			25					30						
Tyr	Glu	Arg	Asp	Glu	Gly	Asp	Lys	Trp	Arg	Asn	Lys	Lys	Phe	Glu	Leu	
		35			40					45						
30	Gly	Leu	Glu	Phe	Pro	Asn	Leu	Pro	Tyr	Tyr	Ile	Asp	Gly	Asp	Val	Lys
		50			55					60						
Leu	Thr	Gln	Ser	Met	Ala	Ile	Ile	Arg	Tyr	Ile	Ala	Asp	Lys	His	Asn	
		65			70					75				80		
Met	Leu	Gly	Gly	Cys	Pro	Lys	Glu	Arg	Ala	Glu	Ile	Ser	Met	Leu	Glu	
		85			90					95						
Gly	Ala	Val	Leu	Asp	Ile	Arg	Tyr	Gly	Val	Ser	Arg	Ile	Ala	Tyr	Ser	
		100			105					110						
Lys	Asp	Phe	Glu	Thr	Leu	Lys	Val	Asp	Phe	Leu	Ser	Lys	Leu	Pro	Glu	
		115			120					125						
35	Met	Leu	Lys	Met	Phe	Glu	Asp	Arg	Leu	Cys	His	Lys	Thr	Tyr	Leu	Asn
		130			135					140						
Gly	Asp	His	Val	Thr	His	Pro	Asp	Phe	Met	Leu	Tyr	Asp	Ala	Leu	Asp	
		145			150					155				160		

Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Trp Asn  
 225 230 235 240  
 Ala Thr Ser His His Thr Arg Pro Thr Pro Gln Leu Pro Arg Gly Pro  
 245 250 255  
 Asn

## (2) INFORMATION FOR SEQ ID NO:224:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 267 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 20 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 25 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 30 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Gly Asp  
 225 230 235 240  
 Val Phe Arg Glu Leu Arg Asp Arg Trp Asn Ala Thr Ser His His Thr  
 245 250 255  
 Arg Pro Thr Pro Gln Leu Pro Arg Gly Pro Asn  
 260 265

## 35 (2) INFORMATION FOR SEQ ID NO:225:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 277 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

5 Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 10 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 15 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 20 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser His  
 225 230 235 240  
 Ser Gly Gly Met Asn Arg Ala Tyr Gly Asp Val Phe Arg Glu Leu Arg  
 245 250 255  
 Asp Arg Trp Asn Ala Thr Ser Ala Ala Thr Arg Pro Thr Pro Gln Leu  
 260 265 270  
 Pro Arg Gly Pro Asn  
 275

(2) INFORMATION FOR SEQ ID NO:226:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 277 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 35 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn

65	70	75	80
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
85	90	95	
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
100	105	110	
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
115	120	125	
5 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn			
130	135	140	
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
145	150	155	160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165	170	175	
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180	185	190	
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195	200	205	
10 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
210	215	220	
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Ala			
225	230	235	240
Arg Asp Ser Gly Pro Ala Glu Asp Gly Ser Arg Ala Val Arg Leu Asn			
245	250	255	
Gly Val Glu Asn Ala Asn Thr Arg Lys Ser Ser Arg Ser Asn Pro Arg			
260	265	270	
15 Gly Arg Arg His Pro			
275			

## (2) INFORMATION FOR SEQ ID NO:227:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 257 amino acids
- (B) TYPE: amino acid
- 20 (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro			
1 5 10 15			
25 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu			
20 25 30			
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu			
35 40 45			
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys			
50 55 60			
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn			
65 70 75 80			
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
30 85 90 95			
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
100 105 110			
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
115 120 125			
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn			
130 135 140			
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
145 150 155 160			
35 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165 170 175			
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180 185 190			
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			

195	200	205
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg		
210	215	220
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Ala		
225	230	235
Arg Asp Ser Gly Pro Ala Glu Asp Gly Ser Arg Ala Val Arg Leu Asn		
245	250	255

5 Gly

(2) INFORMATION FOR SEQ ID NO:228:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 259 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

10

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro			
1	5	10	15
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu			
15	20	25	30
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu			
35	40	45	
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys			
50	55	60	
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn			
65	70	75	80
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
85	90	95	
20 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
100	105	110	
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
115	120	125	
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn			
130	135	140	
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
145	150	155	160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
25 165	170	175	
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180	185	190	
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195	200	205	
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
210	215	220	
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Asp Gly			
225	230	235	240
30 Ser Arg Ala Val Arg Leu Asn Gly Val Glu Asn Ala Asn Thr Arg Lys			
245	250	255	
Ser Ser Arg			

(2) INFORMATION FOR SEQ ID NO:229:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 257 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

35

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 5 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 10 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 15 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Glu Asn  
 225 230 235 240  
 Ala Asn Thr Arg Lys Ser Ser Arg Ser Asn Pro Arg Gly Arg Arg His  
 245 250 255  
 20 Pro

## (2) INFORMATION FOR SEQ ID NO:230:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:

25 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 35 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn

130	135	140
Gly Asp His Val Thr His	Pro Asp Phe Met	Leu Tyr Asp Ala Leu Asp
145	150	155
Val Val Leu Tyr Met Asp Pro Met	Cys Leu Asp Ala Phe Pro	Lys Leu
165	170	175
Val Cys Phe Lys Lys Arg Ile Glu Ala	Ile Pro Gln Ile Asp Lys Tyr	
180	185	190
5 Leu Lys Ser Ser Lys Tyr Ile Ala	Trp Pro Leu Gln Gly Trp Gln Ala	
195	200	205
Thr Phe Gly Gly Asp His Pro Pro	Lys Ser Asp Leu Val Pro Arg	
210	215	220
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala	Ala Ala Ser Glu Asn	
225	230	235
Ala Asn Thr Arg Lys Ser Ser Arg		240
		245

## 10 (2) INFORMATION FOR SEQ ID NO:231:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro	10	15
1	5	
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu	20	25
20		30
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu	35	40
35		45
20 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys	50	55
50		60
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn	65	70
65		75
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu	85	90
85		95
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser	100	105
100		110
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu	115	120
115		125
25 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn	130	135
130		140
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp	145	150
145		155
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro	165	170
165		175
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr	180	185
180		190
30 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala	195	200
195		205
Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg	210	215
210		220
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Arg Lys	225	230
225		235
Ser Ser Arg Ser Asn Pro Arg Gly		240
		245

## 35 (2) INFORMATION FOR SEQ ID NO:232:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 247 amino acids
- (B) TYPE: amino acid

(C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

5 Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 10 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 15 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 20 210 215 220  
 20 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Asn  
 225 230 235 240  
 Pro Arg Gly Arg Arg His Pro  
 245

(2) INFORMATION FOR SEQ ID NO:233:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 249 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

30 Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 35 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110

Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 5 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Thr Arg  
 225 230 235 240  
 10 Lys Ser Ser Arg Ser Asn Pro Arg Gly  
 245

## (2) INFORMATION FOR SEQ ID NO:234:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 277 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- 15 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 20 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 25 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 30 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Thr  
 225 230 235 240  
 35 Pro Pro Ser Arg Glu Ala Tyr Ser Arg Pro Tyr Ser Val Asp Ser Asp  
 245 250 255  
 Ser Asp Thr Asn Ala Lys His Ser Ser His Asn Arg Arg Leu Arg Thr  
 260 265 270  
 Arg Ser Arg Pro Asn

275

## (2) INFORMATION FOR SEQ ID NO:235:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 258 amino acids  
 (B) TYPE: amino acid  
 5 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 10 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 15 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 20 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Thr  
 225 230 235 240  
 25 Pro Pro Ser Arg Glu Ala Tyr Ser Arg Pro Tyr Ser Val Asp Ser Asp  
 245 250 255  
 Ser Asp

## (2) INFORMATION FOR SEQ ID NO:236:

## (i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 259 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

35 Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu

35	40	45	
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys			
50	55	60	
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn			
65	70	75	80
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
85	90	95	
5 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
100	105	110	
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
115	120	125	
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn			
130	135	140	
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
145	150	155	160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165	170	175	
10 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180	185	190	
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195	200	205	
Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
210	215	220	
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Arg			
225	230	235	240
15 Pro Tyr Ser Val Asp Ser Asp Ser Asp Thr Asn Ala Lys His Ser Ser			
245	250	255	
His Asn Arg			

## (2) INFORMATION FOR SEQ ID NO:237:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 257 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

25	1	5	10	15
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro				
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu				
20	25	30		
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu				
35	40	45		
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys				
50	55	60		
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn				
65	70	75	80	
30 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu				
85	90	95		
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser				
100	105	110		
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu				
115	120	125		
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn				
130	135	140		
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp				
145	150	155	160	
35 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu				
165	170	175		
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr				
180	185	190		

Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Thr Asn  
 225 230 235 240  
 Ala Lys His Ser Ser His Asn Arg Arg Leu Arg Thr Arg Ser Arg Pro  
 245 250 255  
 5 Asn

## (2) INFORMATION FOR SEQ ID NO:238:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 247 amino acids
- (B) TYPE: amino acid
- 10 (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 15 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 20 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 25 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Thr Asn  
 30 225 230 235 240  
 Ala Lys His Ser Ser His Asn  
 245

## (2) INFORMATION FOR SEQ ID NO:239:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- 35 (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 5 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 10 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 15 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Ser Ser  
 225 230 235 240  
 His Asn Arg Arg Leu Arg Thr Arg  
 245

20

## (2) INFORMATION FOR SEQ ID NO:240:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

25

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 30 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 35 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp

145	150	155	160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165	170	175	
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180	185	190	
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195	200	205	
5 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
210	215	220	
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ser Arg Arg			
225	230	235	240
Leu Arg Thr Arg Ser Arg Pro Asn			
245			

## (2) INFORMATION FOR SEQ ID NO:241:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 282 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

15	Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro		
1	5	10	15
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu			
20	25	30	
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu			
35	40	45	
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys			
50	55	60	
20 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn			
65	70	75	80
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
85	90	95	
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
100	105	110	
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
115	120	125	
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn			
130	135	140	
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
145	150	155	160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165	170	175	
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180	185	190	
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195	200	205	
30 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
210	215	220	
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Arg Val			
225	230	235	240
Gly Gln Cys Thr Asp Ser Asp Val Arg Arg Pro Trp Ala Arg Ser Cys			
245	250	255	
Ala His Gln Gly Cys Gly Ala Gly Thr Arg Asn Ser His Gly Cys Ile			
260	265	270	
Thr Arg Pro Leu Arg Gln Ala Ser Ala His			
35 275	280		

## (2) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 257 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 10 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 15 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 20 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Arg Val  
 225 230 235 240  
 Gly Gln Cys Thr Asp Ser Asp Val Arg Arg Pro Trp Ala Arg Ser Cys  
 245 250 255  
 Ala

25

(2) INFORMATION FOR SEQ ID NO:243:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 259 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80

Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 10 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Val Arg  
 225 230 235 240  
 Arg Pro Trp Ala Arg Ser Cys Ala His Gln Gly Cys Gly Ala Gly Thr  
 245 250 255  
 Arg Asn Ser

15 (2) INFORMATION FOR SEQ ID NO:244:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 257 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 25 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 30 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 35 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ser Gly Thr

225	230	235	240
Arg Asn Ser His Gly Cys Ile Thr Arg Pro Leu Arg Gln Ala Ser Gln			
245		250	255

His

## (2) INFORMATION FOR SEQ ID NO:245:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro			
1	5	10	15
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu			
20	25		30
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu			
35	40	45	
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys			
50	55	60	
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn			
65	70	75	80
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
85	90		95
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
100	105		110
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
115	120		125
20	Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn		
130	135	140	
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
145	150	155	160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165		170	175
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180	185		190
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195	200		205
25	Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg		
210	215	220	
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Arg Tyr			
225	230	235	240
Lys His Asp Ile Gly Cys Asp Ala Gly Val Asp Lys Lys Ser Ser Ser			
245		250	255
Val Arg Gly Gly Cys Gly Ala His Ser Ser Pro Pro Arg Ala Gly Arg			
260	265		270
30	Gly Pro Arg Gly Thr Met Val Ser Arg Leu		
275	280		

## (2) INFORMATION FOR SEQ ID NO:246:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 262 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 5 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 10 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 145 150 155 160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
 165 170 175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
 180 185 190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
 195 200 205  
 15 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
 210 215 220  
 Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Arg Tyr  
 225 230 235 240  
 Lys His Asp Ile Gly Cys Asp Ala Gly Val Asp Lys Lys Ser Ser Ser  
 245 250 255  
 Val Arg Gly Gly Cys Gly  
 260

20 (2) INFORMATION FOR SEQ ID NO:247:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 264 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

25 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro  
 1 5 10 15  
 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu  
 20 25 30  
 Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu  
 30 35 40 45  
 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
 50 55 60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 65 70 75 80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
 85 90 95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
 100 105 110  
 35 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
 115 120 125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
 130 135 140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp

145	150	155	160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165	170	175	
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180	185	190	
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195	200	205	
5 Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
210	215	220	
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Gly Cys			
225	230	235	240
Asp Ala Gly Val Asp Lys Lys Ser Ser Val Arg Gly Gly Cys Gly			
245	250	255	
Ala His Ser Ser Pro Pro Arg Ala			
260			

## 10 (2) INFORMATION FOR SEQ ID NO:248:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 259 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro			
1 5 10 15			
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu			
20 25 30			
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu			
35 40 45			
20 Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys			
50 55 60			
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn			
65 70 75 80			
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu			
85 90 95			
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser			
100 105 110			
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu			
115 120 125			
25 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn			
130 135 140			
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp			
145 150 155 160			
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu			
165 170 175			
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr			
180 185 190			
30 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala			
195 200 205			
Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg			
210 215 220			
Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser Gly Ala			
225 230 235 240			
His Ser Ser Pro Pro Arg Ala Gly Arg Gly Pro Arg Gly Thr Met Val			
245 250 255			
Ser Arg Leu			

35

## (2) INFORMATION FOR SEQ ID NO:249:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

Ser Gly Ser Pro Pro Cys Cys Ser Trp Gly Arg Phe Met Gln Gly  
 1 5 10 15  
 Gly Leu Phe Gly Gly Arg Thr Asp Gly Cys Gly Ala His Arg Asn Arg  
 20 25 30  
 Thr Ser Ala Ser Leu Glu Pro Pro Ser Ser Asp Tyr  
 35 40

10 (2) INFORMATION FOR SEQ ID NO:250:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

Ser His Ser Gly Gly Met Asn Arg Ala Tyr Gly Asp Val Phe Arg Glu  
 1 5 10 15  
 Leu Arg Asp Arg Trp Asn Ala Thr Ser His His Thr Arg Pro Thr Pro  
 20 25 30  
 Gln Leu Pro Arg Gly Pro Asn Ser  
 35 40

20 (2) INFORMATION FOR SEQ ID NO:251:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

25 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

Asp Thr Asn Ala Lys His Ser Ser His Asn Arg Arg Leu Arg Thr Arg  
 1 5 10 15  
 Ser Arg Pro Asn Gly  
 20

30 (2) INFORMATION FOR SEQ ID NO:252:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

Cys Gly Ala Gly Thr Arg Asn Ser His Gly Cys Ile Thr Arg Pro Leu  
 1 5 10 15

Arg Gln Ala Ser Ala His Gly  
20

## (2) INFORMATION FOR SEQ ID NO:253:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:

(A) NAME/KEY: Modified Site

(B) LOCATION: 1

(D) OTHER INFORMATION: "Xaa=Ser or Thr"

(A) NAME/KEY: Modified Site

(B) LOCATION: 3

(D) OTHER INFORMATION: "Xaa=Arg or Lys"

(A) NAME/KEY: Modified Site

(B) LOCATION: 4

(D) OTHER INFORMATION: "Xaa=Lys or Arg"

(A) NAME/KEY: Modified Site

(B) LOCATION: 6

(D) OTHER INFORMATION: "Xaa=Ser or Leu"

(A) NAME/KEY: Modified Site

(B) LOCATION: 7

(D) OTHER INFORMATION: "Xaa=Arg, Ile, Val or Ser"

(A) NAME/KEY: Modified Site

(B) LOCATION: 8

(D) OTHER INFORMATION: "Xaa=Ser, Tyr, Phe or His"

(A) NAME/KEY: Modified Site

(B) LOCATION: 10

(D) OTHER INFORMATION: "Xaa=Phe, His or Arg"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

25 Xaa Thr Xaa Xaa Ser Xaa Xaa Xaa Asn Xaa Arg  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:254:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:

(A) NAME/KEY: Modified Site

(B) LOCATION: 2

(D) OTHER INFORMATION: "Xaa=Ser, Ala or Gly"

(A) NAME/KEY: Modified Site

(B) LOCATION: 4

(D) OTHER INFORMATION: "Xaa=Val or Gln"

(A) NAME/KEY: Modified Site  
 (B) LOCATION: 7  
 (D) OTHER INFORMATION: "Xaa=Pro, Gly or Ser"

(A) NAME/KEY: Modified Site  
 (B) LOCATION: 8  
 (D) OTHER INFORMATION: "Xaa=Trp or Tyr"

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

Asp Xaa Asp Xaa Arg Arg Xaa Xaa  
 1 5

(2) INFORMATION FOR SEQ ID NO:255:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide  
 (ix) FEATURE:

15 (A) NAME/KEY: Modified Site  
 (B) LOCATION: 7  
 (D) OTHER INFORMATION: "Xaa=Ala or Phe"

(A) NAME/KEY: Modified Site  
 (B) LOCATION: 8  
 (D) OTHER INFORMATION: "Xaa=Arg or His"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

20 Val Arg Ser Gly Cys Gly Xaa Xaa Ser Ser  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:256:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

Asn Thr Arg Lys Ser Ser Arg Ser Asn Pro Arg  
 1 5 10

30 (2) INFORMATION FOR SEQ ID NO:257:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: unknown

35 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

Ser Thr Lys Arg Ser Leu Ile Tyr Asn His Arg

1

5

10

## (2) INFORMATION FOR SEQ ID NO:258:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

5

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

Ser Thr Gly Arg Lys Val Phe Asn Arg Arg  
1 5 10

10

## (2) INFORMATION FOR SEQ ID NO:259:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

15

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

Thr Asn Ala Lys His Ser Ser His Asn Arg Arg  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:260:

20

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

25

Asp Ser Asp Val Arg Arg Pro Trp  
1 5

## (2) INFORMATION FOR SEQ ID NO:261:

30

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

Ala Ala Asp Gln Arg Arg Gly Trp  
1 5

35

## (2) INFORMATION FOR SEQ ID NO:262:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

5 Asp Gly Arg Gly Gly Arg Ser Tyr  
1 5

(2) INFORMATION FOR SEQ ID NO:263:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids

- (B) TYPE: amino acid

10 (C) STRANDEDNESS:

- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

Arg Val Arg Ser  
1

15 (2) INFORMATION FOR SEQ ID NO:264:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids

- (B) TYPE: amino acid

- (C) STRANDEDNESS:

- (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

Ser Val Arg Ser Gly Cys Gly Phe Arg Gly Ser Ser  
1 5 10

(2) INFORMATION FOR SEQ ID NO:265:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids

- (B) TYPE: amino acid

- (C) STRANDEDNESS:

- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

30 Ser Val Arg Gly Gly Cys Gly Ala His Ser Ser  
1 5 10

WHAT IS CLAIMED IS:

1. A purified protein which specifically binds to a gastro-intestinal tract receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI.

2. A protein which binds specifically to a gastro-intestinal transport receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, in which the protein comprises an amino acid sequence selected from the group consisting of SEQ ID NOS:1-55 or a binding portion thereof.

3. A protein which binds specifically to a gastro-intestinal transport receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, in which the amino acid sequence of the protein is selected from the group consisting of SEQ ID NOS:1-55, or a binding portion thereof.

20 4. The protein of claim 2 which comprises the amino acid sequence substantially as set forth in: SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 30, SEQ ID NO: 43, SEQ ID NO: 46, or SEQ ID NO: 52, or a binding portion thereof.

25

5. The protein of claim 3, the amino acid sequence of which consists of the amino acid sequence substantially as set forth in: SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 30, SEQ ID NO: 43, SEQ ID NO: 46, or SEQ ID NO: 52, or a binding portion thereof.

30 6. A protein of not more than 50 amino acids in length which specifically binds to a gastro-intestinal transport receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, in which the protein includes, positioned anywhere along its sequence, the contiguous amino

acid sequence of: Xaa<sub>1</sub> Thr Xaa<sub>2</sub> Xaa<sub>3</sub> Ser Xaa<sub>4</sub> Xaa<sub>5</sub> Xaa<sub>6</sub> Asn Xaa, Arg (SEQ ID NO:253), where Xaa<sub>1</sub> is Ser or Thr; Xaa<sub>2</sub> is Arg or Lys; Xaa<sub>3</sub> is Lys or Arg; Xaa<sub>4</sub> is Ser or Leu; Xaa<sub>5</sub> is Arg, Ile, Val, or Ser; Xaa<sub>6</sub> is Ser, Tyr, Phe, or His; and Xaa,  
5 is Pro, His or Arg.

7. The protein of claim 6 which is not more than 40 amino acids in length.

10 8. The protein of claim 6 which is not more than 30 amino acids in length.

9. The protein of claim 6 which is not more than 20 amino acids in length.

15 10. A protein of not more than 50 amino acids in length which specifically binds to a gastro-intestinal transport receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, in which the protein includes,  
20 positioned anywhere along its sequence, the contiguous amino acid sequence of: Asp Xaa<sub>1</sub> Asp Xaa<sub>2</sub> Arg Arg Xaa<sub>3</sub> Xaa<sub>4</sub> (SEQ ID NO:254) where Xaa<sub>1</sub> is Ser, Ala, or Gly; Xaa<sub>2</sub> is Val or Gln; Xaa<sub>3</sub> is Pro, Gly, or Ser; and Xaa<sub>4</sub> is Trp or Tyr.

25 11. The protein of claim 10 which is not more than 40 amino acids in length.

12. The protein of claim 10 which is not more than 30 amino acids in length.

30 13. The protein of claim 10 which is not more than 20 amino acids in length.

14. A protein of not more than 50 amino acids in  
35 length which specifically binds to a gastro-intestinal transport receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, in which the protein includes,

positioned anywhere along its sequence, the contiguous amino acid sequence of: Val Arg Ser Gly Cys Gly Xaa<sub>1</sub> Xaa<sub>2</sub> Ser Ser (SEQ ID NO:255), where Xaa<sub>1</sub> is Ala or Phe; and Xaa<sub>2</sub> is Arg or His.

5

15. The protein of claim 14 which is not more than 40 amino acids in length.

10 16. The protein of claim 14 which is not more than 30 amino acids in length.

17. The protein of claim 14 which is not more than 20 amino acids in length.

15 18. A protein of not more than 50 amino acids in length which specifically binds to a gastro-intestinal transport receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, in which the protein includes, positioned anywhere along its sequence, the contiguous amino acid sequence of: NTRKSSRSNPR (SEQ ID NO:256) or STKRSLIYNHR (SEQ ID NO:257) or STGRKVFNRR (SEQ ID NO:258) or TNAKHSSHNR (SEQ ID NO:259).

19. A protein of not more than 50 amino acids in length which specifically binds to a gastro-intestinal transport receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, in which the protein includes, positioned anywhere along its sequence, the contiguous amino acid sequence of: DSDVRRPW (SEQ ID NO:260) or AADQRRGW (SEQ 30 ID NO:261) or DGRGGRSY (SEQ ID NO:262).

20. A protein of not more than 50 amino acids in length which specifically binds to a gastro-intestinal transport receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, in which the protein includes, positioned anywhere along its sequence, the contiguous amino

acid sequence of: RVRS (SEQ ID NO:263) or SVRSGCGFRGSS (SEQ ID NO:264) or SVRGGCGAHSS (SEQ ID NO:265).

21. The protein of claim 1, 2, 3, 6, 10, 14, 18, 5 19, or 20 which is purified.

22. A composition comprising the protein of claim 1, 2, 3, 6, 10, 14, 18, 19, or 20, bound to a material comprising an active agent, said active agent being of value 10 in the treatment of a mammalian disease or disorder.

23. The composition of claim 22 in which the active agent is a drug.

15 24. The composition of claim 22 in which the material is a particle containing the active agent.

25. The composition of claim 22 in which the material is a slow-release device containing the drug.

20 26. The composition of claim 22 in which the protein is covalently or noncovalently bound to the material.

27. A composition comprising a chimeric protein 25 bound to a material comprising an active agent, in which the chimeric protein comprises a sequence selected from the group consisting of SEQ ID NOS:1-55 or a binding portion thereof fused via a covalent bond to an amino acid sequence of a second protein, in which the active agent is of value in the 30 treatment of a mammalian disease or disorder.

28. A composition comprising the protein of claim 1, 2, 3, 6, 10, 14, 18, 19, or 20 covalently bound to a particle containing a drug.

35

29. A composition comprising the protein of claim 1, 2, 3, 6, 10, 14, 18, 19, or 20 covalently bound to a drug.

30. The composition of claim 22 which facilitates the transport of the active agent through human or animal gastro-intestinal tissue.

5 31. A method of delivering an active agent *in vivo* comprising administering to a subject a purified composition of claim 22.

10 32. A method of delivering a drug to a subject comprising administering to the subject a purified composition of claim 30.

15 33. A method of delivering a drug to a subject comprising administering to the subject a purified composition of claim 31.

34. The method according to claim 31 in which the administering is oral.

20 35. The method according to claim 31 in which the active agent is a drug.

36. The method according to claim 31 in which the subject is a human.

25 37. The method according to claim 35 in which the subject is a human.

38. The method according to claim 31 in which said 30 composition facilitates the transport of the active agent through human or animal gastro-intestinal tissue.

39. The method according to claim 33 in which the administering is oral.

40. A pharmaceutical composition comprising the composition of claim 22 in a pharmaceutically acceptable carrier suitable for use in humans *in vivo*.

5 41. A chimeric protein comprising at least 6 contiguous amino acids of an amino acid sequence selected from the group consisting of SEQ ID NOS:1-55, that specifically bind to a gastro-intestinal tract receptor, fused via a covalent bond to an amino acid sequence of a  
10 second protein.

42. An antibody which is capable of immunospecifically binding the protein of claim 2, 3, 6, 10, 14, 18, 19 or 20.

15

43. A molecule comprising a fragment of the antibody of claim 42, which fragment is capable of immunospecifically binding said protein.

20

44. A purified derivative of the protein of claim 1 or 2, which displays one or more functional activities of said protein.

25

45. The derivative of claim 44 which is able to be bound by an antibody directed against said protein.

46. A fragment of the protein of claim 2 comprising a domain of said protein.

30

47. A fragment of the protein of claim 3 comprising a domain of said protein.

35

48. A nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:110-163.

49. A nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:55-109.

5 50. An isolated nucleic acid comprising a nucleotide sequence encoding the protein of claim 1.

51. A nucleic acid comprising a nucleotide sequence encoding the protein of claim 2, 3, 6, 10, 14, 18, 10 19 or 20.

52. The nucleic acid of claim 51 which is a DNA.

53. The nucleic acid of claim 48 or 49 which is 15 isolated.

54. The nucleic acid of claim 51 which is isolated.

20 55. An isolated nucleic acid comprising a nucleotide sequence complementary to the nucleotide sequence of claim 57.

56. An isolated nucleic acid comprising a 25 nucleotide sequence encoding a fragment of the protein of claim 1, 2, or 3, which fragments bind to said gastrointestinal tract receptor.

57. A nucleic acid comprising a nucleotide 30 sequence encoding the chimeric protein of claim 41.

58. A nucleic acid comprising a nucleotide sequence encoding the fragment of claim 47.

35 59. The nucleic acid of claim 57 which is isolated.

60. The nucleic acid of claim 58 which is isolated.

61. A recombinant cell containing the nucleic acid 5 of claim 48, 49 or 50.

62. A recombinant cell containing the nucleic acid of claim 51.

10 63. A recombinant cell containing the nucleic acid of claim 57.

64. A method of producing a protein comprising growing a recombinant cell containing the nucleic acid of 15 claim 48, 49 or 50 such that the encoded protein is expressed by the cell, and recovering the expressed protein.

65. A method of producing a protein comprising growing a recombinant cell containing the nucleic acid of 20 claim 51 such that the encoded protein is expressed by the cell, and recovering the expressed protein.

66. A method of producing a protein comprising growing a recombinant cell containing the nucleic acid of 25 claim 57 such that the encoded protein is expressed by the cell, and recovering the expressed protein.

67. The product of the process of claim 64.

30 68. The product of the process of claim 65.

69. The product of the process of claim 66.

70. A pharmaceutical composition comprising a 35 therapeutically effective amount of a composition comprising the protein of claim 1, 2, 3, 6, 10, 14, 18, 19, or 20; and a pharmaceutically acceptable carrier.

71. The chimeric protein of claim 41 in which said second protein is a drug.

72. A nucleic acid comprising a nucleotide sequence encoding the protein of claim 71.

73. A pharmaceutical composition comprising a therapeutically effective amount of the protein of claim 71, and a pharmaceutically acceptable carrier.

10

74. A pharmaceutical composition comprising a therapeutically effective amount of the nucleic acid of claim 78.

15

75. A method of delivering a drug to a subject comprising administering to the subject a therapeutically effective amount of the pharmaceutical composition of claim 80.

20

76. A method of treating or preventing a disease or disorder comprising administering to a subject in which such treatment or prevention is desired a therapeutically effective amount of the composition of claim 23.

25

77. A method of treating or preventing a disease or disorder comprising administering to a subject in which such treatment or prevention is desired a therapeutically effective amount of the composition of claim 28.

30

78. A method of treating or preventing a disease or disorder comprising administering to a subject in which such treatment or prevention is desired a therapeutically effective amount of the composition of claim 29.

35

79. The method according to claim 76 in which the disease or disorder is selected from the group consisting of:

hypertension, diabetes, osteoporosis, hemophilia, anemia, cancer, migraines, and angina pectoris.

80. The method according to claim 76 in which the  
5 subject is a human.

81. A composition comprising the protein of claim  
1, 2, 3, 6, 10, 14, 18, 19, 20, or 46 wherein the protein is  
coated onto or absorbed onto or covalently bonded to the  
10 surface of a nano- or microparticle.

82. A nano- or microparticle formed from the  
protein of claim 1, 2, 3, 6, 10, 14, 18, 19, 20, or 46.

15 83. The composition of claim 87, wherein the nano-  
or microparticle is a drug-loaded or drug-encapsulating nano-  
or microparticle.

84. A method of detecting or measuring the level  
20 of a gastro-intestinal tract receptor in a sample, comprising  
contacting a sample suspected of containing a gastro-  
intestinal tract receptor with the protein of claim 1, 2, 3,  
6, 10, 14, 18, 19, 20, or 46 under conditions conducive to  
binding between the protein and any of said receptor in said  
25 sample, and detecting or measuring any of said binding that  
occurs, in which the detected or measured amount of binding  
indicates the presence or amount of the receptor in the  
sample.

30 85. A method of identifying a molecule that  
specifically binds to a ligand selected from the group  
consisting of the protein of claim 1, 2, 3, 6, 10, 14, 18, or  
19, a fragment of said protein comprising a domain of the  
protein, and a nucleic acid encoding said protein or  
35 fragment, comprising

(a) contacting said ligand with a plurality of molecules under conditions conducive to binding between said ligand and the molecules; and

(b) identifying a molecule within said plurality 5 that specifically binds to said ligand.

86. An isolated nucleic acid encoding a fragment of a gastro-intestinal tract receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, or encoding a 10 chimeric protein comprising said fragment, said fragment consisting essentially of the extracellular domain of the receptor.

87. A cell containing and capable of expressing a 15 recombinant nucleic acid encoding a fragment of a gastro-intestinal tract receptor selected from the group consisting of HPT1, hPEPT1, D2H, and hSI, or encoding a chimeric protein comprising said fragment, said fragment consisting essentially of the extracellular domain of the receptor.

20

88. The cell of claim 87 which contains an expression vector comprising a nucleotide sequence encoding said fragment operably linked to a heterologous promoter.

25

89. A method for identifying a molecule that specifically binds to a gastro-intestinal tract receptor comprising contacting a fragment of the receptor, or a chimeric protein comprising said fragment, with a plurality of test molecules under conditions conducive to binding 30 between said fragment or protein and the molecules, and identifying a molecule within said plurality that specifically binds to said fragment or protein, in which the fragments consist essentially of the extracellular domain of the receptor.

35

90. The composition of claim 22 for use as a medicament.

91. The composition of claim 28 for use as a medicament.

92. The composition of claim 29 for use as a  
5 medicament.

93. The composition of claim 81 for use as a medicament.

10 94. The composition of claim 23 in which the drug is insulin or leuprolide.

95. The composition of claim 24 in which the active agent is insulin or leuprolide.

15

96. The composition of claim 25 in which the drug is insulin or leuprolide.

97. The composition of claim 28 in which the drug  
20 is insulin or leuprolide.

25

30

35

1/48

20	40	60
MGMSKSHSFFGYPLSIFFIV VNEFCERFSYYGMRAILILY FTNFISWDDNLSTAIYHTFV		
80	100	120
ALCYLTPILGALIADSWLGK FKTIVSLSIVYTIGQAVTSV SSINDLTDHNHDGTPDSLKV		
140	160	180
HVVLISLIGLALIALGTGGIK PCVSAFGGDQFEEGQEKQRN RFFSIFYLAINAGSLLSTII		
200	220	240
TPMLRVQQCGIHSKQACYPL AFGVPAALMAVALIVFVLGS GMYKKFKPQGNIMGKVAKCI		
260	280	300
GFAIKNRFRHRSKAFPKREH WLDWAKEKYDERLISQIKMV TRVMFLYIPLPMFWALFDQQ		
320	340	360
GSRWTLQATTMSGKIGALEI QPDQMQTVNAILIVIMVPIF DAVLYPLIAKCGFNFTSLKK		
380	400	420
MAVGMVLASMAFVVAIIVQV EIDKTPVFPKGNEVQIKVL NIGNNTMNISLPGEMVTLGP		
440	460	480
MSQTNAFMTFDVNKLTRINI SSPGSPVTAVTDDFKQGQRH TLLWAPNHYQVVKDGLNQK		
500	520	540
PEKGENGIRFVNTFNELITI TMSGKVYANISSYNASTYQF FPSGIKGFTISSTEIPPQCQ		
560	580	600
PNFNTFYLEFGSAYTYIVQR KNDSCPEVKFEDISANTVN MALQIPQYFLLTCGEVVFSV		
620	640	660
TGLEFSYSQAPSNMKSVLQA GWLLTVAVGNIIVLIVAGAG QFSKQWAEYILFAALLLVVC		
680	700	708
VIFAIMARFYTYINPAEIEA QFDEDEKKNRLEKSNPYFMS GANSQKQM		

FIG.1

1 gaattccgtc tcgaccactg aatggaagaa aaggacttt aaccaccatt ttgtgactta  
 61 cagaaaggaa tttgaataaa gaaaactatg atacttcagg cccatcttca ctccctgtgt  
 M I L Q A H L H S L C  
 121 cttcttatgc tttatttggc aactggatat ggccaagagg ggaagtttag tggaccctg  
 L L M L Y L A T G Y G Q E G K F S G P L  
 181 aaacccatga cattttctat ttatgaaggc caagaaccga gtcaaattat attccagttt  
 K P M T F S I Y E G Q E P S Q I I F Q F  
 241 aaggccaatc ctcctgtgt gactttgaa ctaactgggg agacagacaa catatttg  
 K A N P P A V T F E L T G E T D N I F V  
 301 atagaacggg agggacttct gtattacaac agagccttgg acagggaaac aagatctact  
 I E R E G L L Y Y N R A L D R E T R S T  
 361 cacaatctcc agttgcagc cctggacgct aatggaatta tagtggaggg tccagtcct  
 H N L Q V A A L D A N G I I V E G P V P  
 421 atcaccatag aagtgaagga catcaacgac aatcgaccac cgtttctcca gtcaaagtac  
 I T I E V K D I N D N R P T F L Q S K Y  
 481 gaaggctcag taaggcagaa ctctcgccca gaaaagccct tcttgatgt caatgccaca  
 E G S V R Q N S R P G K P F L Y V N A T  
 541 gacctggatg atccggccac tcccaatggc cagtttattt accagattgt catccagctt  
 D L D D P A T P N G Q L Y Y Q I V I Q L  
 601 cccatgatca acaatgtcat gtacttcag atcaacaaca aaacgggagc catctctt  
 P M I N N V M Y F Q I N N K T G A I S L  
 661 acccgagagg gatctcagga attgaatcct gctaagaatc cttcctataa tctggtgatc  
 T R E G S Q E L N P A K N P S Y N L V I  
 721 tcagtgaagg acatgggagg ccagagttag aattccttca gtgataccac atctgtggat  
 S V K D M G G Q S E N S F S D T T S V D  
 781 atcatagtga cagagaatat ttggaaagca cccaaacctg tggagatggt gggaaactca  
 I I V T E N I W K A P K P V E M V E N S  
 841 actgatcctc accccatcaa aatcaactcag gtgcgttgg atgatcccgg tgcacaatat  
 T D P H P I K I T Q V R W N D P G A Q Y  
 901 tccttagttg acaaagagaa gctgccaaga ttccatattt caattgacca ggaaggagat  
 S L V D K E K L P R F P F S I D Q E G D  
 961 atttacgtga ctcaagccctt ggaccgagaa gaaaaggatg catatgtttt ttatgcagtt  
 I Y V T Q P L D R E E K D A Y V F Y A V  
 1021 gcaaaggatg agtacggaaa accacttca tatccgttgg aaattcatgt aaaagttaaa  
 A K D E Y G K P L S Y P L E I H V K V K  
 1081 gatattaatg ataatccacc tacatgtccg tcaccagtaa ccgtatttga ggtccaggag  
 D I N D N P P T C P S P V T V F E V Q E  
 1141 aatgaacgac tggtaacag tatcgggacc cttactgcac atgacaggaa tgaagaaaat  
 N E R L G N S I G T L T A H D R D E E N  
 1201 actgccaaca gtttctaaa ctacaggatt gtggagcaaa ctcccaaact tcccatggat  
 T A N S F L N Y R I V E Q T P K L P M D

1261 ggactcttcc taatccaaac ctatgctgga atgttacagt tagctaaaca gtccttgaag  
     G L F L I Q T Y A G M L Q L A K Q S L K  
 1321 aagcaagata ctcctcagta caacttaacg atagaggtgt ctgacaaaga tttcaagacc  
     K Q D T P Q Y N L T I E V S D K D F K T  
 1381 ctgttttg tgcaaataaa cgttattgtat atcaatgatc agatccccat ctttgaaaaa  
     L C F V Q I N V I D I N D Q I P I F E K  
 1441 tcagattatg gaaacctgac tcttgctgaa gacacaaaca ttgggtccac catcttaacc  
     S D Y G N L T L A E D T N I G S T I L T  
 1501 atccaggcca ctgatgctga tgagccattt actgggagtt ctaaaattct gtatcatatc  
     I Q A T D A D E P F T G S S K I L Y H I  
 1561 ataaagggag acagtgggg acgcctgggg gttgacacag atccccatac caacaccgga  
     I K G D S E G R L G V D T D P H T N T G  
 1621 tatgtcataa ttaaaaagcc tcttgattt gaaacagcag ctgtttccaa cattgttttc  
     Y V I I K K P L D F E T A A V S N I V F  
 1681 aaagcagaaa atcctgagcc tctagtgttt ggtgtgaagt acaatgcaag ttctttgcc  
     K A E N P E P L V F G V K Y N A S S F A  
 1741 aagttcacgc ttattgtgac agatgtgaat gaagcacctc aattttccca acacgtattc  
     K F T L I V T D V N E A P Q F S Q H V F  
 1801 caagcgaaag tcagtggagga tggtagctata ggcactaaag tggcaatgt gactgccaag  
     Q A K V S E D V A I G T K V G N V T A K  
 1861 gatccagaag gtctggacat aagctattca ctgagggggg acacaagagg ttggcttaaa  
     D P E G L D I S Y S L R G D T R G W L K  
 1921 attgaccacg tgactggta gatcttagt gtggctccat tggacagaga agccgaaagt  
     I D H V T G E I F S V A P L D R E A G S  
 1981 ccatatcggt tacaagtggt ggccacagaa gtaggggggt cttccttaag ctctgtgtca  
     P Y R V Q V V A T E V G G S S L S S V S  
 2041 gagttccacc tgatccttat ggtatgtgaat gacaaccctc ccaggctagc caaggactac  
     E F H L I L M D V N D N P P R L A K D Y  
 2101 acgggcttgt tcttctgcca tcccctcagt gcacctggaa gtctcatttt cgaggctact  
     T G L F F C H P L S A P G S L I F E A T  
 2161 gatgatgatc agacttatt tcgggtccc cattttacat tttccctcggt cagtggaaac  
     D D D Q H L F R G P H F T F S L G S G S  
 2221 ttacaaaacg actggaaatg ttccaaaatc aatggtactc atgcccact gtcaccagg  
     L Q N D W E V S K I N G T H A R L S T R  
 2281 cacacagact ttgaggagag ggcgtatgtc gtcttgcattcc gcatcaatga tgggggtcggt  
     H T D F E E R A Y V V L I R I N D G G R  
 2341 ccacccttgg aaggcattgt ttcttacca gttacattct gcagttgtgt ggaagggaaatg  
     P P L E G I V S L P V T F C S C V E G S  
 2401 tggccggc cagcaggta ccagactggg atacccactg tggcatggc agttggata  
     C F R P A G H Q T G I P T V G M A V G I

2461 ctgctgacca cccttctgg tattggata atttagcag ttgtgttat ccgcataaag  
L L T T L L V I G I I L A V V F I R I K  
2521 aaggataaaag gcaaagataa tggtaaagt gctcaagcat ctgaagtcaa acctctgaga  
K D K G K D N V E S A Q A S E V K P L R  
2581 agctgaattt gaaaaggaat gtttgeattt atatagcaag tgctattca gcaacaacca  
S

2641 tctcatccta ttactttca tctaacgtgc attataattt tttaaacaga tattccctct  
2701 tgtccttaa tatttgctaa atattcttt tttgaggtgg agtcttgctc tgtcgccag  
2761 gctggagtagt agtgggtgtga tcccagctca ctgcaacctc cgcctcctgg gttcacatga  
2821 ttctcctgcc tcagcttcct aagtagctgg gtttacaggc acccaccacc atgcccagct  
2881 aattttgtat ttttaatag agacggggtt tcgcccattt gccaggatgg tcttgeactc  
2941 ctgacgtcaa gtgatctgcc tgccttggtc tcccaataca ggcataacc actgcaccca  
3001 cctacttaga tatttcatgt gctatagaca ttagagagat tttcatttt tccatgacat  
3061 ttttcctctc tgcaaatggc ttagctactt gtgttttcc ctttggggc aagacagact  
3121 cattaaatat tctgtacatt ttttctttat caaggagata tatcagtgtt gtctcataga  
3181 actgcctgga ttccatttat gtttttctg attccatcct gtgtccccctt catccttgac  
3241 tccttggta ttcaactgaa tttcaaacat ttgtcagaga agaaaaaaagt gaggactcag  
3301 gaaaaataaaa taaaataaaag aacagcctt tgcggcccg aattc

FIG.2C

20	40	60
MARKKFSGLEISLIVLFVIV	TIIAIALIVVLATKTPAVDE	ISDSTSTPATTRVTTNPSDS
80	100	120
GKCPNVLNDPVNVRINCIPE	QFPTEGICAQRGCCWRPWND	SLIPWCFFVDNHGYNVQDMT
140	160	180
TTSIGVEAKLNRIPSPLTFG	NDINSVLFTTQNQTPNRFRF	KITDPNNRRYEVPHQYVKEF
200	220	240
TGPTVSDTLYDVKVAQNPFS	IQVIRKSNGKTLFDTSIGPL	VYSDQYLQISARLPSDYIYG
260	280	300
IGEQVHKRFRHDL SWKTWPI	FTRDQLPGDNNNNLYGHQTF	FMCIEDTSGKSFGVFLMNSN
320	340	360
AMEIFIQPTPIVTYRVTGGI	LDFYILLGDTPEQVVQQYQQ	LVGLPAMPAYWNLGFQLSRW
380	400	420
NYKSLDVVKEVVRNNREAGI	PFDTQVTIDYMEDKKDFTY	DQVAFNGLPQFVQDLHDHGQ
440	460	480
KYVIILDPAISIGRRANGTT	YATYERGNTQHWWINESDGS	TPIIGEVWPGLTVYPDFTNP
500	520	540
NCIDWWANEC SIFHQEVQYD	GLWIDMNEVSSFIQGSTKGC	NVNKLNYPPFTPDILDKLMY
560	580	600
SKTICMDAVQNWGKQYDVHS	LYGYSMAIATEQAVQKVFPN	KRSFILTRSTFAGSGRHAAH
620	640	660
WLGDNNTASWEQMEWSITGML	EFSLFGIPLVGADICGFVAE	TTEELCRRWMQLGAFYPFSR
680	700	720
NHNSDGYEHQDPAFFGQNSL	LVKSSRQYLTIRYTLPLFLY	TLFYKAHVGETVARPVHL
740	760	780
FYEDTNSWIEDTEFLWGPAL	LITPVLKQGADTVSAYIPDA	IWYDYESGAKRPWRKQRVDM
800	820	840
YL PADKIGLHLRGGYI IPIQ	EPDVTTTASRKNPLGLIVAL	GENNTAKGDFFWDDGETKDT
860	880	900
IQONGNYI LYTFSVSNNTLDI	VCTHSSYQEGTTLAFQTVKI	LGLTDSVTEVRVAENNQPMN
920	940	960
AHSNFTYDASNQVLLIADLK	LNLGRNFSVQWNQIFSENER	FNCYPDADLATEQKCTQRC
980	1000	1020
VWRTGSSL SKAPECYFPRQD	NSYSVNSARYSSMGITADLQ	LNTANARIKLPSPDPISTLRV
1040	1060	1080
EVKYHKNDMLQFKIYDPQKK	RYEVPVPLNIPTTPISTYED	RLYDVEIKENPFGIQIRRRS
1100	1120	1140
SGRVIWDSWLPGFAFNDQFI	QISTRLPSEYIYGFGEV	AFKRDLNWNTWGMFTRDQPP
1160	1180	1200
GYKLNSYGFHPYYMALEEEG	NAHGVFLNSNAMDVTQPT	PALTYRTVGGILDYMFMLGP
1220	1240	1260
TPQVATKQYHEVIGHPVMPA	YWALGFQLCRYGYANTSEVR	ELYDAMVAANIPYDVQYTDI

6/48

1280	1300	1320
DYMERQLDFTIGEAFQDLPQ	FVDKIRGEGMRYIIILDPAI	SGNETKTPAFERGQQNDVF
1340	1360	1380
VKWPNTNDICWAKVWPDLPN	ITIDKTLTEDEAVNASRAHV	AFPDFFRTSTAEEWWAREIVD
1400	1420	1440
FYNEKMKFDGLWIDMNEPSS	FVNGTTNQCRNDELNYPPY	FPELTKRTDGLHFRTICMEA
1460	1480	1500
EQILSDGTSVLHYDVHNLYG	WSQMKPTHDALQKTTGKRG	VISRSTYPTSGRWGGHWLGD
1520	1540	1560
NYARWDNMDKSIIGMMEFSL	FGISYTGADICGFFNNSEYH	LCTRWMQLGAFYPYSRNHNI
1580	1600	1620
ANTRRQDPASWNETFAEMSR	NILNIRYTLLPYFYTQMHEI	HANGGTIVRPLLHEFFDEKP
1640	1660	1680
TWDIFKQFLWGPAPMVTPL	EPYVQTVNAYVPNARWFDYH	TGKDIGHVRGQFQTFNASYDT
1700	1720	1740
INLHVRGGHILPCQEPAQNT	FYSRQKHMKLIVAADDNQMA	QGSLFWDDGESIDTYERDLY
1760	1780	1800
LSVQFNLNQTTLTSTILKRG	YINKSETRLGSLHVWGKGTT	PVAVTLTYNGNKNSLPFNE
1820	1827	
DTTNMILRIDLTTHNVTLEE	PIEINWS	

FIG.3B

1 gccttactgc aggaaggcac tccgaagaca taagtcggtg agacatggct gaagataaaa  
M A E D K  
61 gcaagagaga ctccatcgag atgagtatga agggatgcc acaaacaac gggtttgc  
S K R D S I E M S M K G C Q T N N G F V  
121 ataatgaaga cattctggag cagaccccg atccaggcag ctcaacagac aaccta  
H N E D I L E Q T P D P G S S T D N L K  
181 acagcaccag gggcatccct ggctcccagg agcccgactt caagggcg  
H S T R G I L G S Q E P D F K G V Q P Y  
241 cggggatgcc caaggaggc tggtccagt tctctggcca ggcccgtac cgcata  
A G M P K E V L F Q F S G Q A R Y R I P  
301 gggagatccct ttctggc acagtggctt ctgtgtgg  
R E I L F W L T V A S V L V L I A A T I  
361 ccatcattgc cctctctcca aagtgcctag actgggtggca ggagggccc atgtacc  
A I I A L S P K C L D W W Q E G P M Y Q  
421 tctacccaag gtcttcaag gacagtaaca agatggaa cggagatctg aaaggatt  
I Y P R S F K D S N K D G N G D L K G I  
481 aagataaaact ggactacatc acagctttaa atataaaaac tggttggatt acttcattt  
Q D K L D Y I T A L N I K T V W I T S F  
541 ataaatcgtc ccttaaagat ttca  
Y K S S L K D F R Y G V E D F R E V D P  
601 ttttggAAC gatggaaat tttgagaatc tggttgcagc catacatgat aaaggatt  
I F G T M E D F E N L V A A I H D K G L  
661 aattaaatcat cgatttcata ccaaaccaca cgagtataa acatatttgg tttcaattga  
K L I I D F I P N H T S D K H I W F Q L  
721 gtcggacacg gacaggaaaa tatactgatt attatatctg gcatgactgt acccatgaaa  
S R T R T G K Y T D Y Y I W H D C T H E  
781 atggcaaaac cattccaccc aacaactggt taagtgtgt  
N G K T I P P N N W L S V Y G N S S W H  
841 ttgacgaagt gcgaaaccaa tggttatttc atcagtttat gaaagagcaa cctgatt  
F D E V R N Q C Y F H Q F M K E Q P D L  
901 atttccgcaa tcctgatgtt caagaagaaa taaaagaaaat ttacggttc tggctcacaa  
N F R N P D V Q E E I K E I L R F W L T  
961 agggtgttga tggtttagt ttggatgctg taaaattcct cctagaagca agcac  
K G V D G F S L D A V K F L L E A K H L

FIG.4A

1021 gagatgagat ccaagtaaat aagacccaaa tcccggacac ggtcacacaa tactcgagc  
     R D E I Q V N K T Q I P D T V T Q Y S E  
 1081 tgtaccatga ctaccaccacc acgcaggtgg gaatgcacga cattgtccgc agcttcggc  
     L Y H D F T T T Q V G M H D I V R S F R  
 1141 agaccatgga ccaatacagc acggagcccg gcagatacag gttcatgggg actgaagcct  
     Q T M D Q Y S T E P G R Y R F M G T E A  
 1201 atgcagagag tattgacagg accgtatgt actatggatt gccatttatac caagaagctg  
     Y A E S I D R T V M Y Y G L P F I Q E A  
 1261 atttccatt caacaattac ctacgatgc tagacactgt ttctggaaac agcgtgtatg  
     D F P F N N Y L S M L D T V S G N S V Y  
 1321 agtttatcac atcctggatg gaaaacatgc cagaaggaaa atggcctaac tggatgattg  
     E V I T S W M E N M P E G K W P N W M I  
 1381 gtggaccaga cagttcacgg ctgacttcgc gtttggggaa tcagtatgtc aacgtatga  
     G G P D S S R L T S R L G N Q Y V N V M  
 1441 acatgcttct tttcacactc cctggaaactc ctataactt ctagggagaa gaaattggaa  
     N M L L F T L P G T P I T Y Y G E E I G  
 1501 tggaaatat tgtagccgca aatctcaatg aaagctatgattatattacc ttgcgtcaaa  
     M G N I V A A N L N E S Y D I N T L R S  
 1561 agtcaccaat gcagtggac aatagttcaa atgctggttt ttctgaagct agtaacacct  
     K S P M Q W D N S S N A G F S E A S N T  
 1621 gtttacctac caattcagat taccacactg tgaatgttga tgtccaaaag actcagccca  
     W L P T N S D Y H T V N V D V Q K T Q P  
 1681 gatcggttt gaagttatat caagattaa gtctacttca tgccaatgag ctactcctca  
     R S A L K L Y Q D L S L L H A N E L L L  
 1741 acaggggctg gtttgccat ttgaggaatg acagccacta tggatgtac acaagagagc  
     N R G W F C H L R N D S H Y V V Y T R E  
 1801 tggatggcat cgacagaatc tttatcggtt ttctgaattt tggagaatca acactgttaa  
     L D G I D R I F I V V L N F G E S T L L  
 1861 atctacataa tatgatttcg ggcctcccg ctaaaataag aataaggta agtaccaatt  
     N L H N M I S G L P A K I R I R L S T N  
 1921 ctggcaca aaggcagtaaa gttgatacaa gtggcatttt tctggacaag ggagagggac  
     S A D K G S K V D T S G I F L D K G E G  
 1981 tcatcttgc acacaacacg aagaatctcc ttcatcgcca aacagcttc agagatagat  
     L I F E H N T K N L L H R Q T A F R D R  
 2041 gctttgttgc caatcgagca tgctattcca gtgtactgaa catactgtat acctcggtt  
     C F V S N R A C Y S S V L N I L Y T S C  
 2101 aggcacctt atgaagagat gaagacactg gcatttcagt gggattgtaa gcatttgtaa  
 2161 tagcttcatg tacagcatgc tgcttggta acaatcatta attcttcgat attctgttag  
 2221 cttgaatgta accgctttaa gaaaggttct caaatgttt gaaaaaaata aaatgtttaa  
 2281 aagt

## EXPRESSION OF PHAGE INSERTS AS GST FUSION

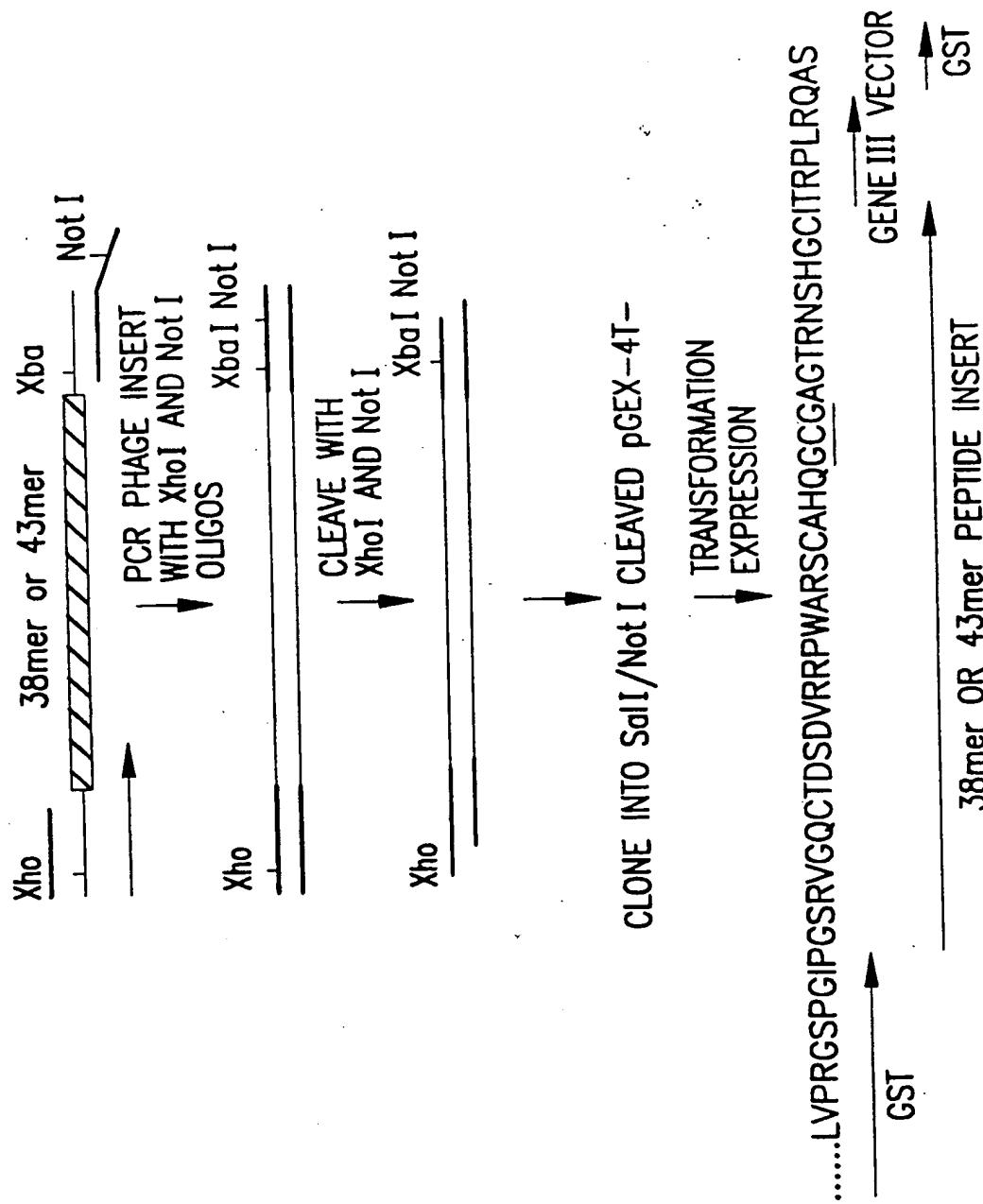


FIG. 5A

10/48

P31	1	10	20	30	Clone #
	SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRRHP				
	SARDSGPAEDGSRAVRLNG				101
	DGSRAVRLNGVENANTRKSSR				102
	ENANTRKSSRSNPRGRRHP				103
	TRKSSRSNPRG				119
Pax2	1	10	20	30	Clone #
	STPPSREAYSRPYSVDSDSDTNAKHSSHNRRRLTRSRPN				
	STPPSREAYSRPYSVDSDSD				104
	SRPYSVDSDSDTNAKHSSHNR				105
	TNAKHSSHNRRRLTRSRPN				106
DCX8	1	10	20	30	Clone #
	RYKHDIGCDAGVDKSSSVRGCGAHSSPPRAGRGPRTMVSRL				
	RYKHDIGCDAGVDKSSSVRGCG				107
	GCDAGVDKSSSVRGCGAHSSPPRA				108
	GAHSSPPRAGRGPRTMVSRL				109

FIG.5B

11/48

P31	1	10	20	30	Clone #
	SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRRHP				
	ENANTRKSSRSNPRGRRHP				103
	ENANTRKSSR				110
	TRKSSRSNPRG				119
	RKSSRSNPRG				111
	SNPRGRRHP				112
Pax2	1	10	20	30	Clone #
	STPPSREAYSRPYSVDSDSDTNAKHSSHNRRRLRTRSRPN				
	TNAKHSSHNRRRLRTRSRPN				106
	TNAKHSSHNN				113
	SSHNRRRLRTR				114
	RRLRTRSRPN				115
SNi10	1	10	20	30	Clone#
	RVGQCTDSDVRRPWARSCAHQGCGAGTRNSHGCITRPLRQASAH				
	RVGQCTDSDVRRPWARSCA				116
	VRRPWARSCAHQGCGAGTRNS				117
	GTRNSHGCITRPLRQASAH				118

FIG.5C

12/48

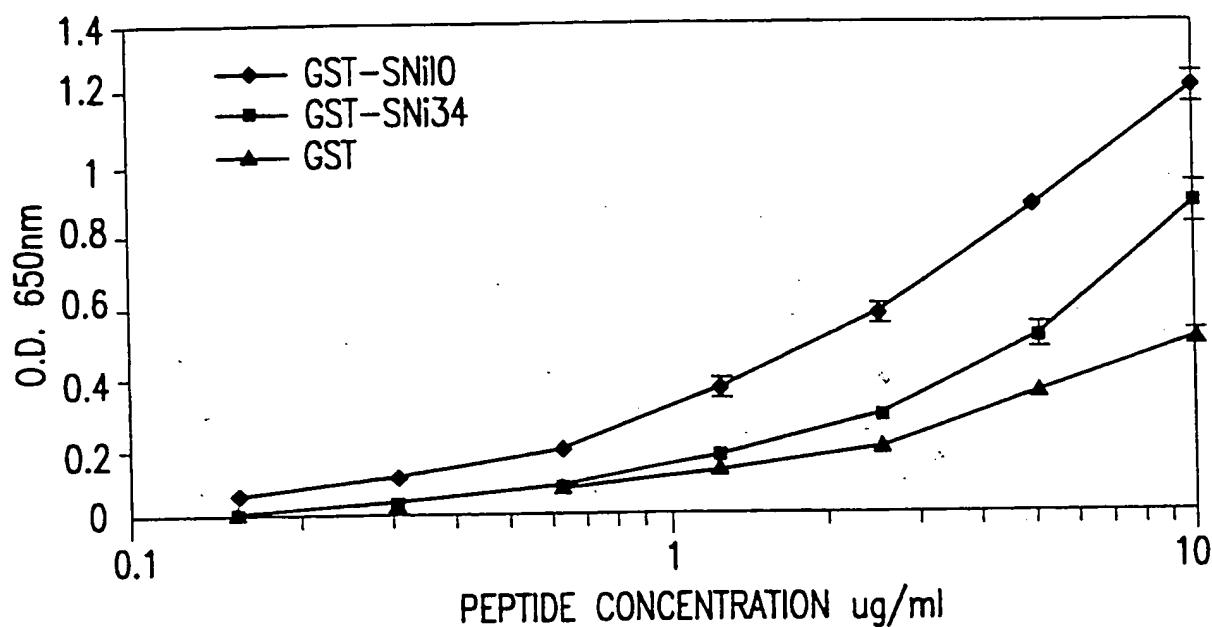


FIG.6A

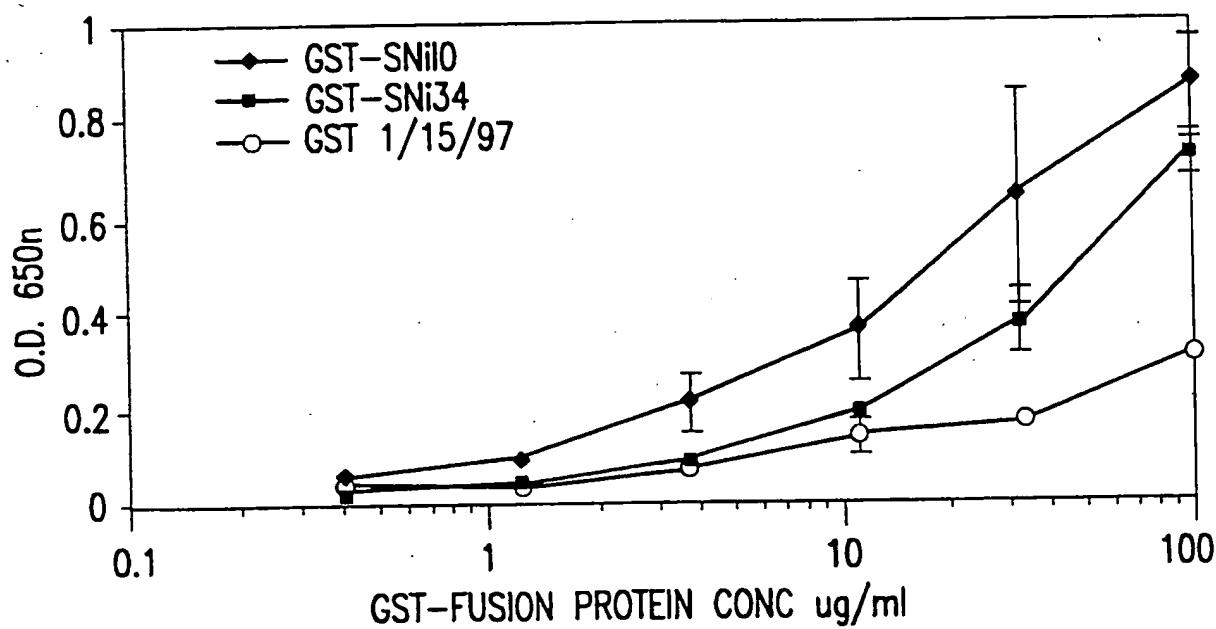


FIG.6B

13/48

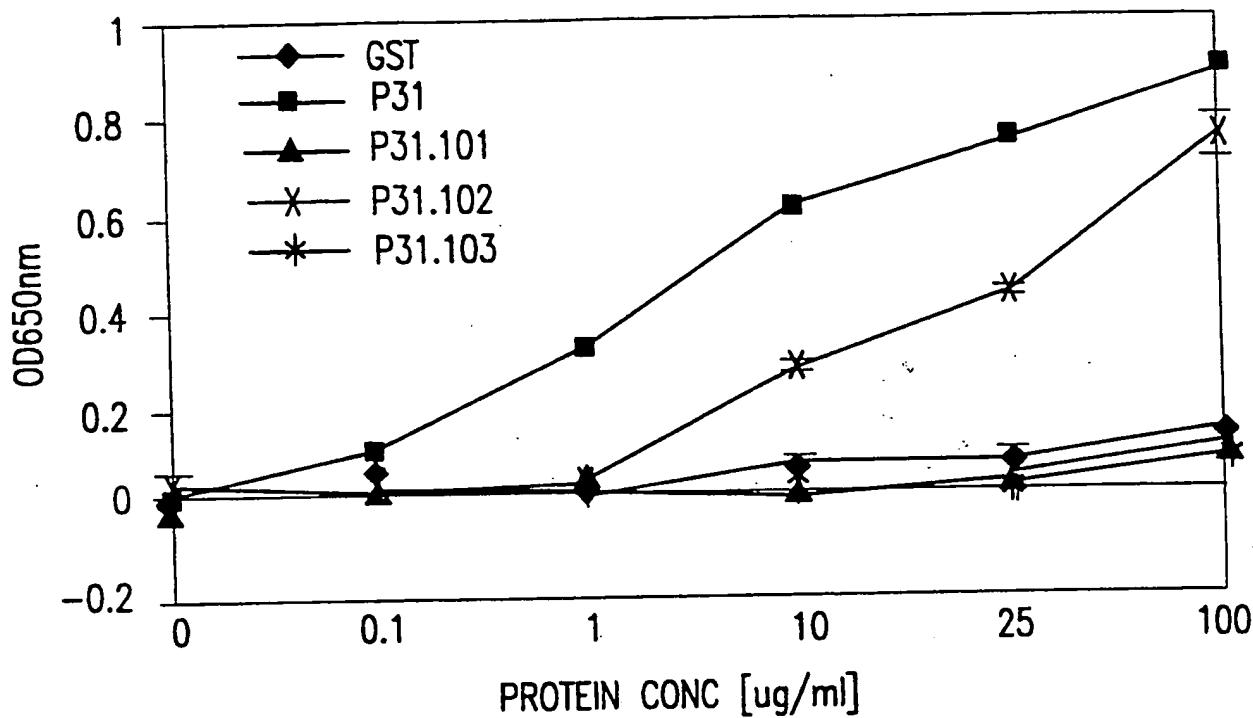


FIG. 7A

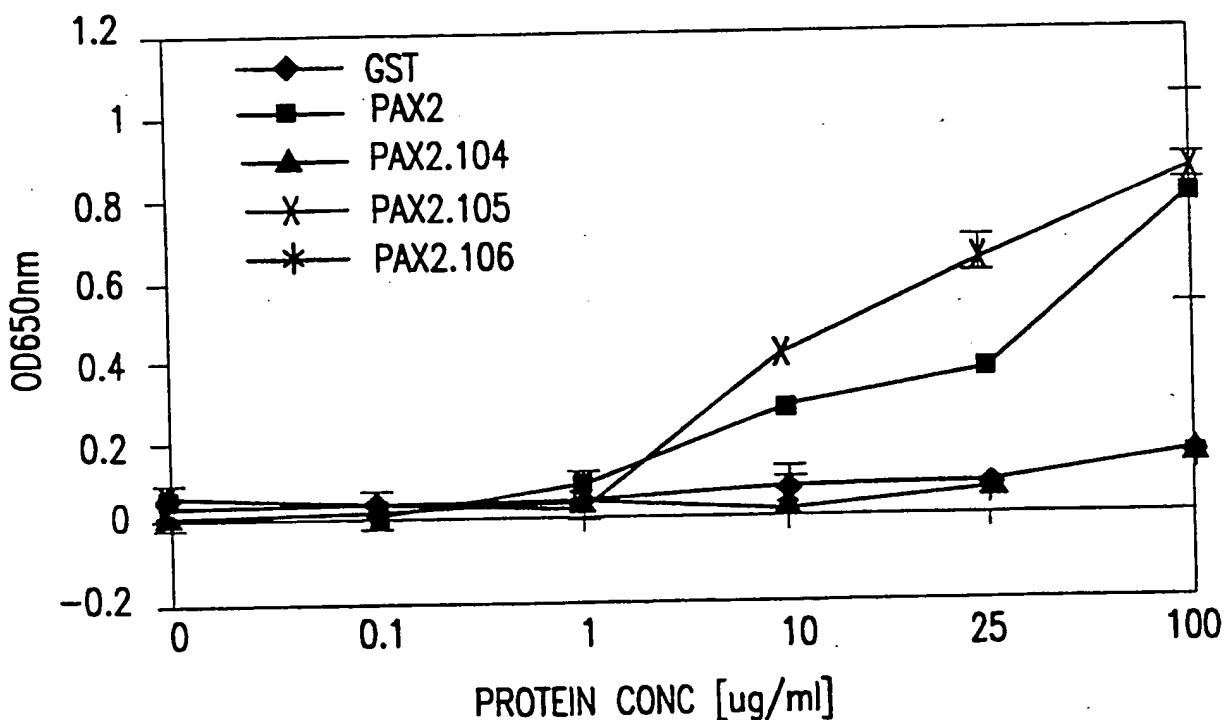


FIG. 7B

14/48

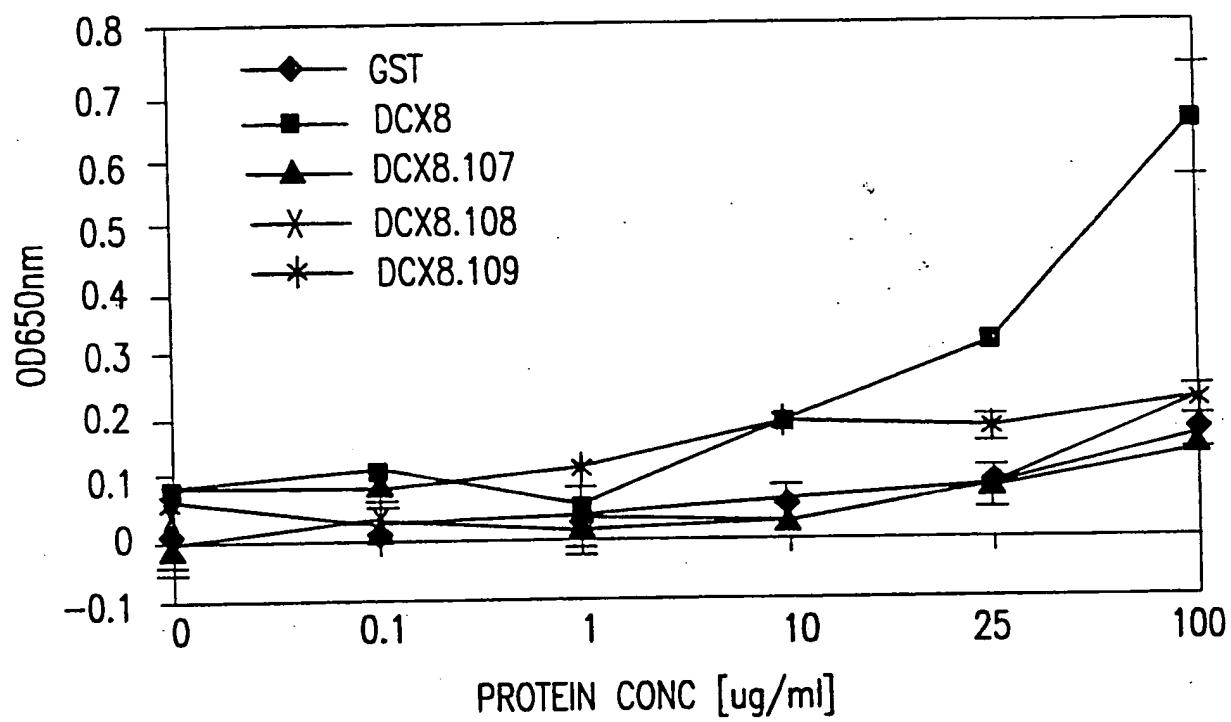


FIG. 7C

15/48

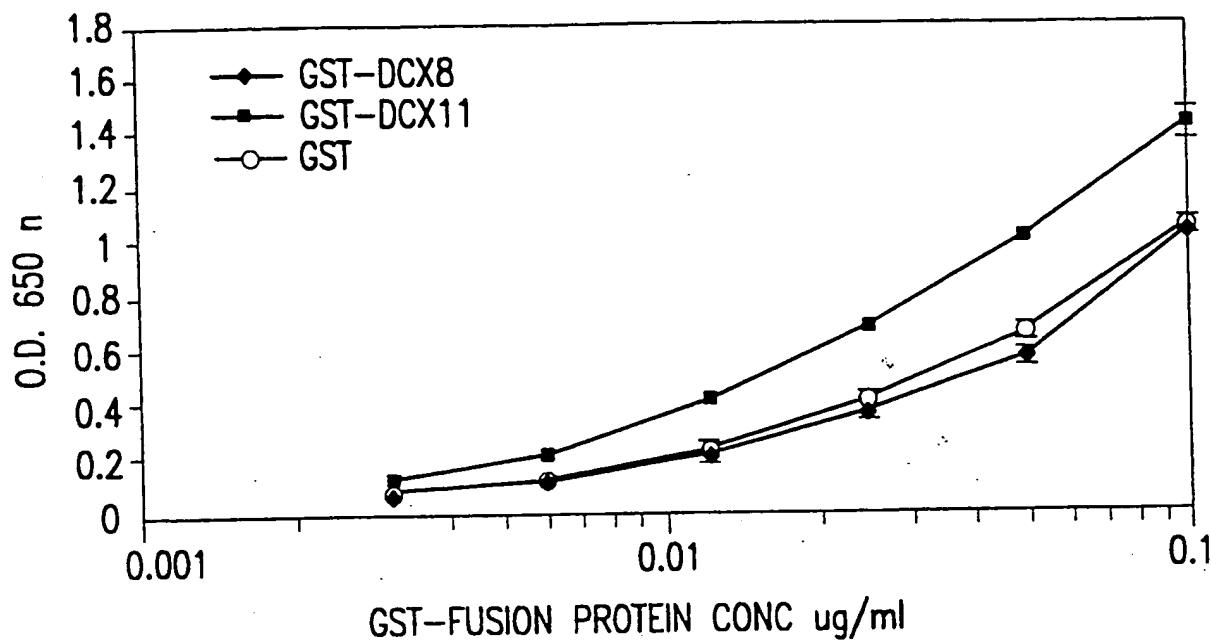


FIG.7D

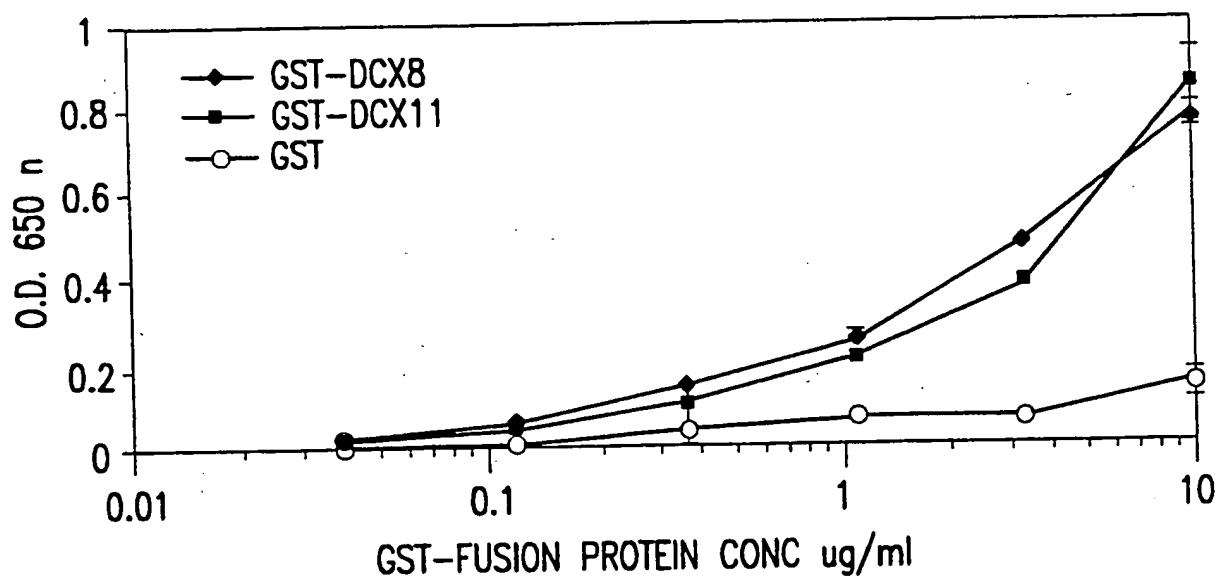


FIG.7E

16/48

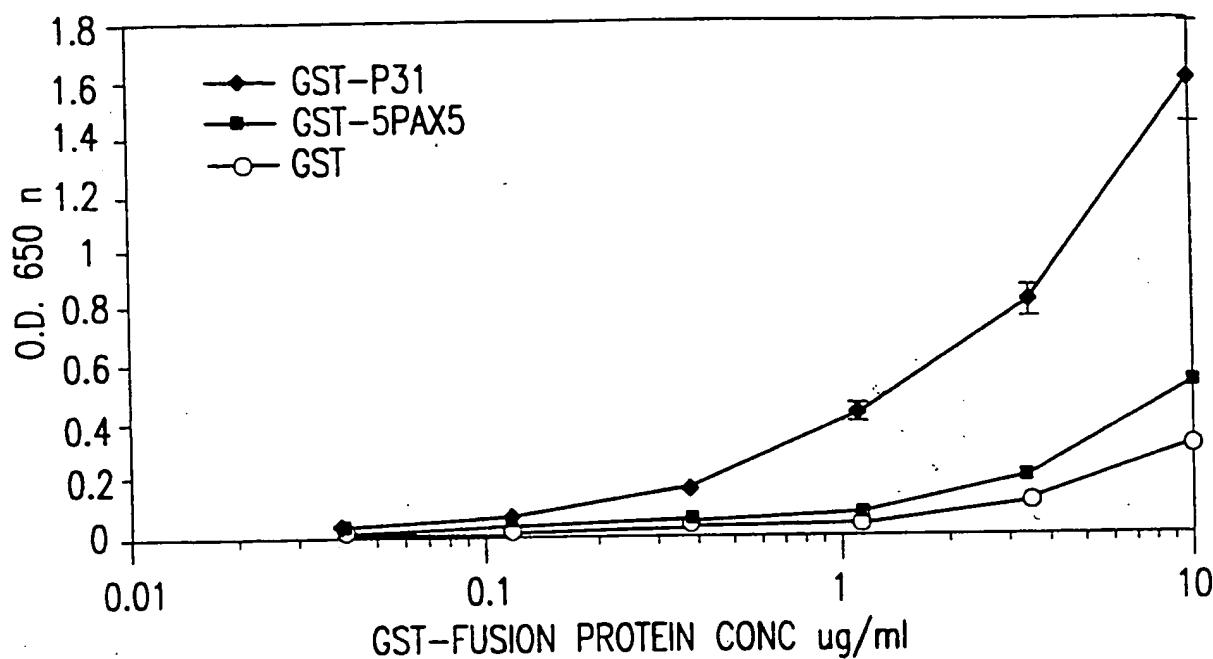


FIG.7F

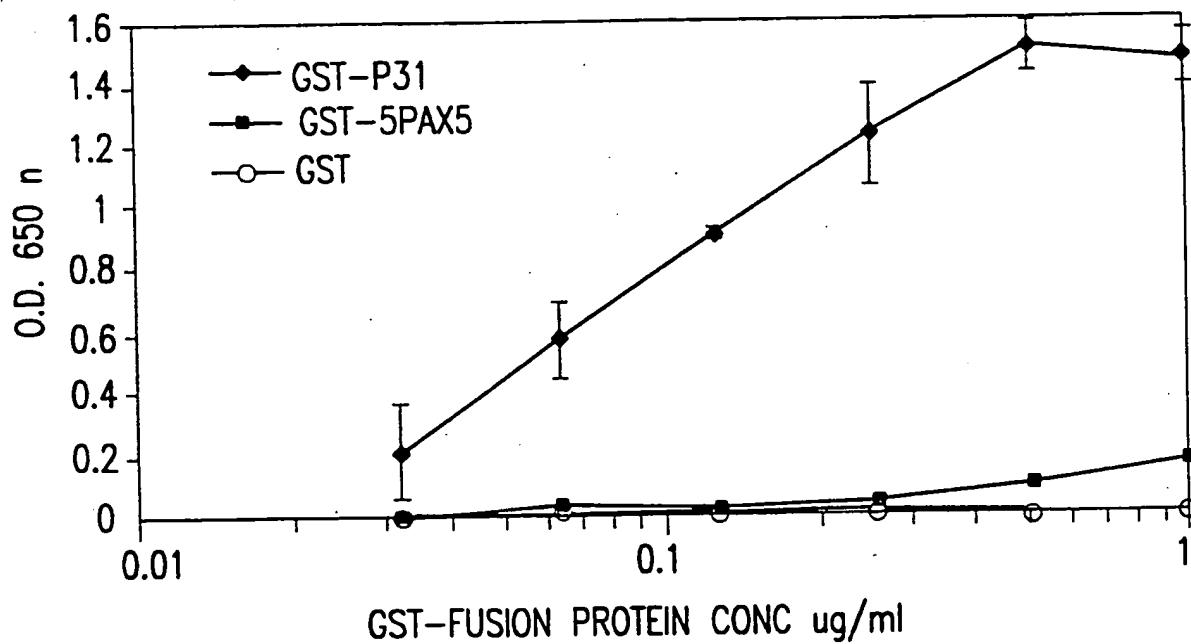


FIG.7G

17/48

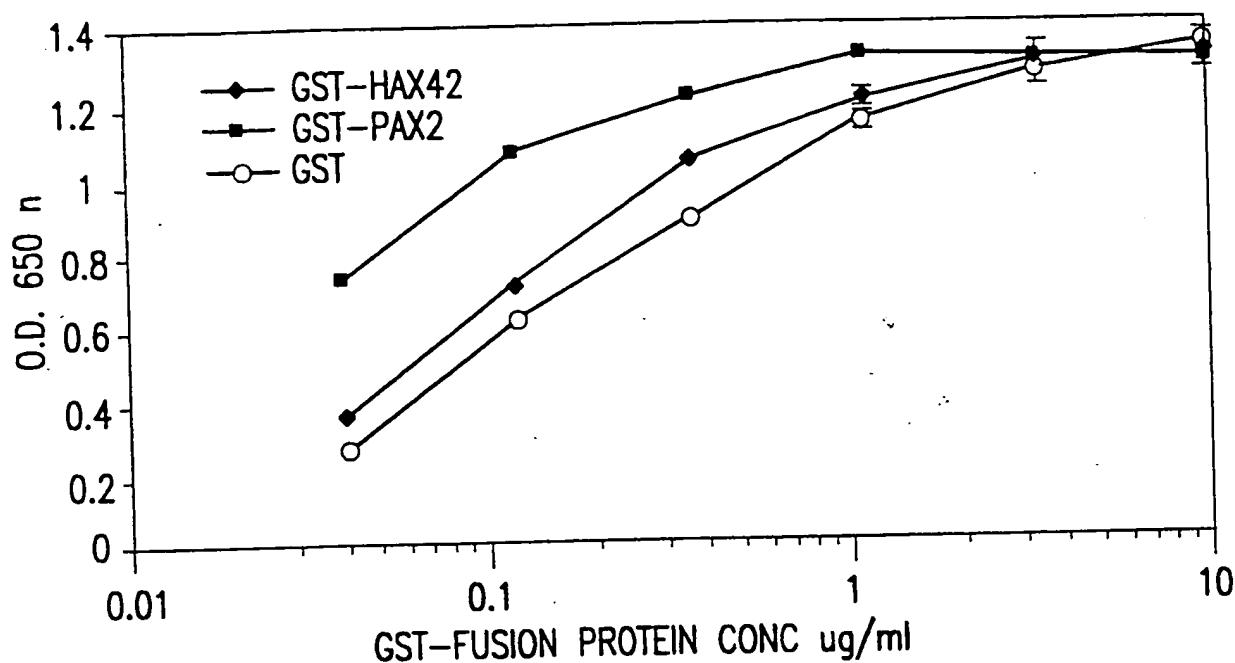


FIG.7H

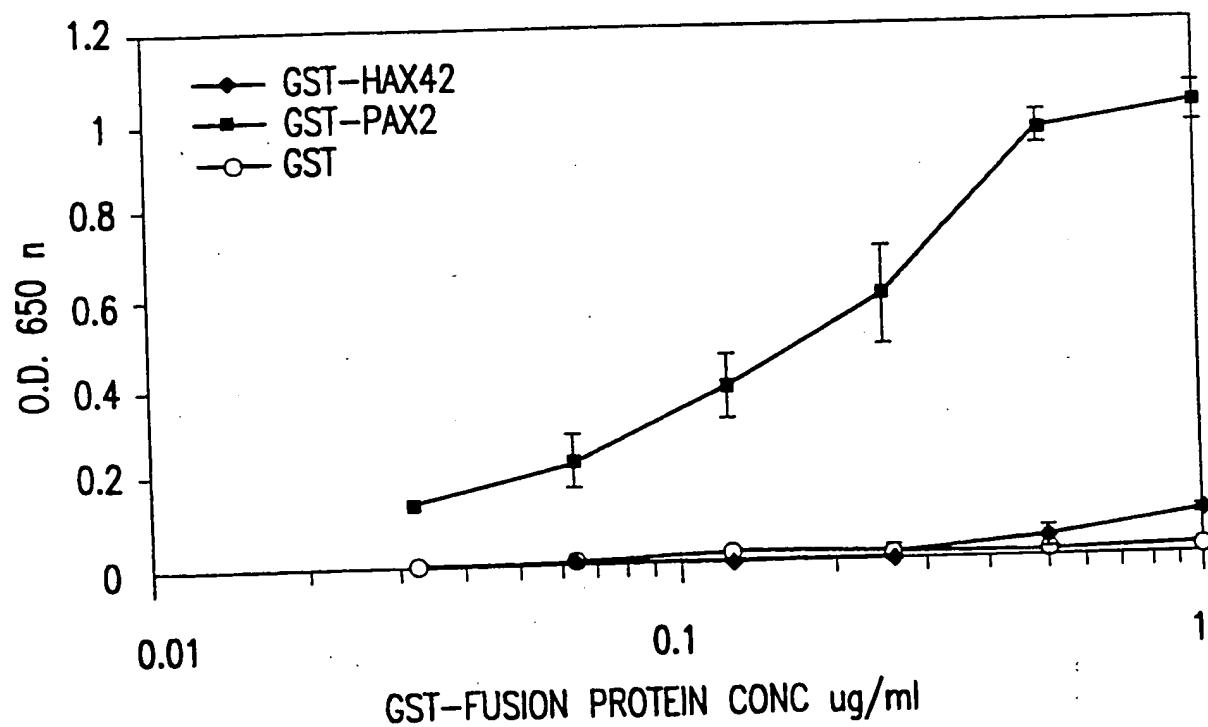


FIG.7I

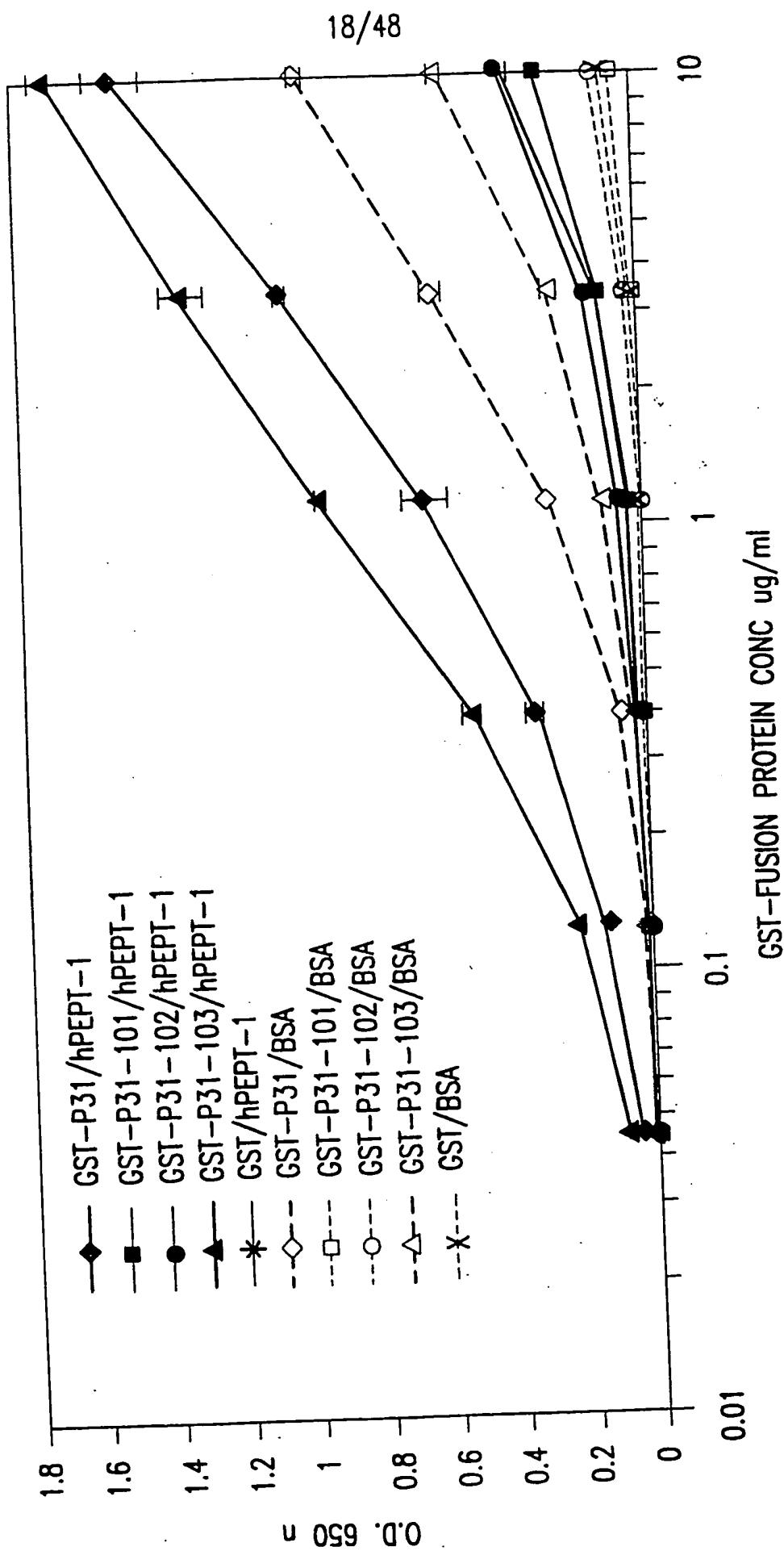


FIG. 7J

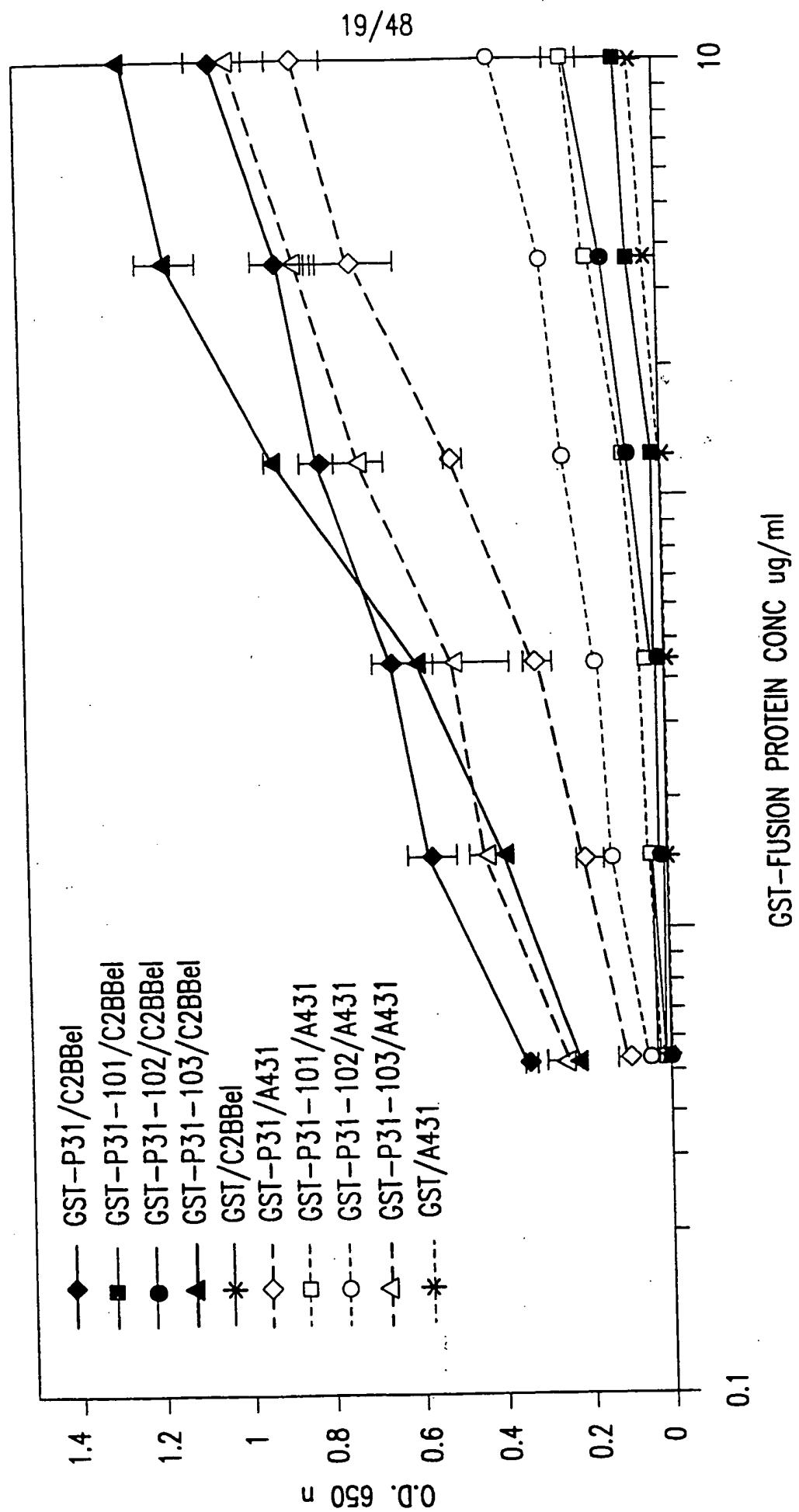


FIG. 7K

20/48

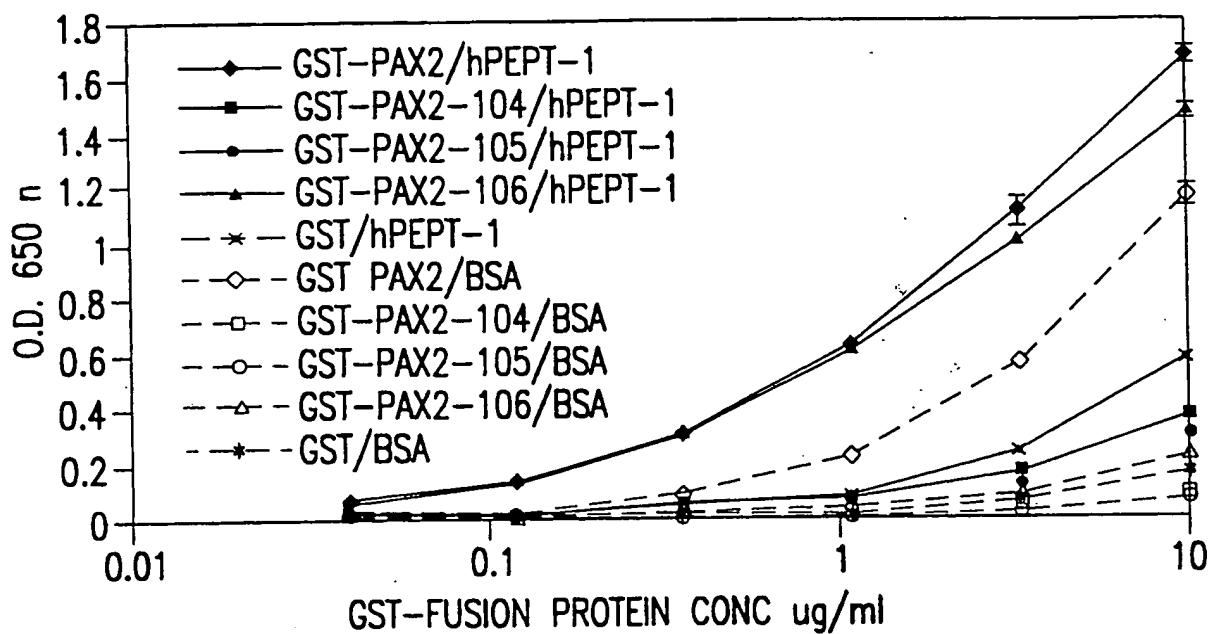


FIG. 7L

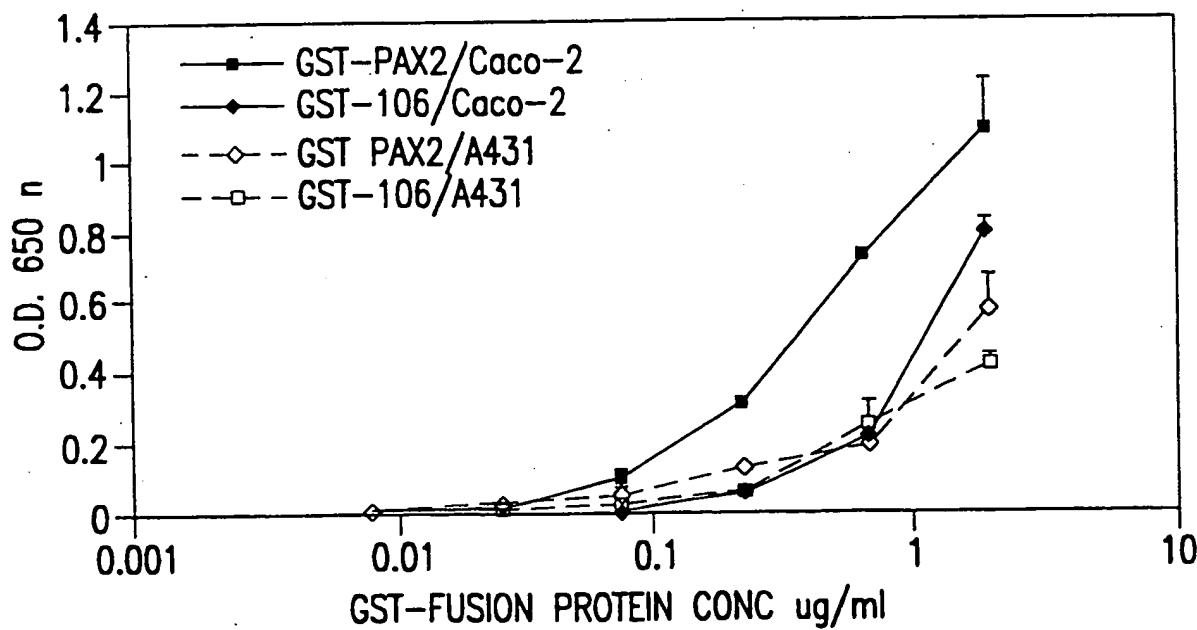


FIG. 7M

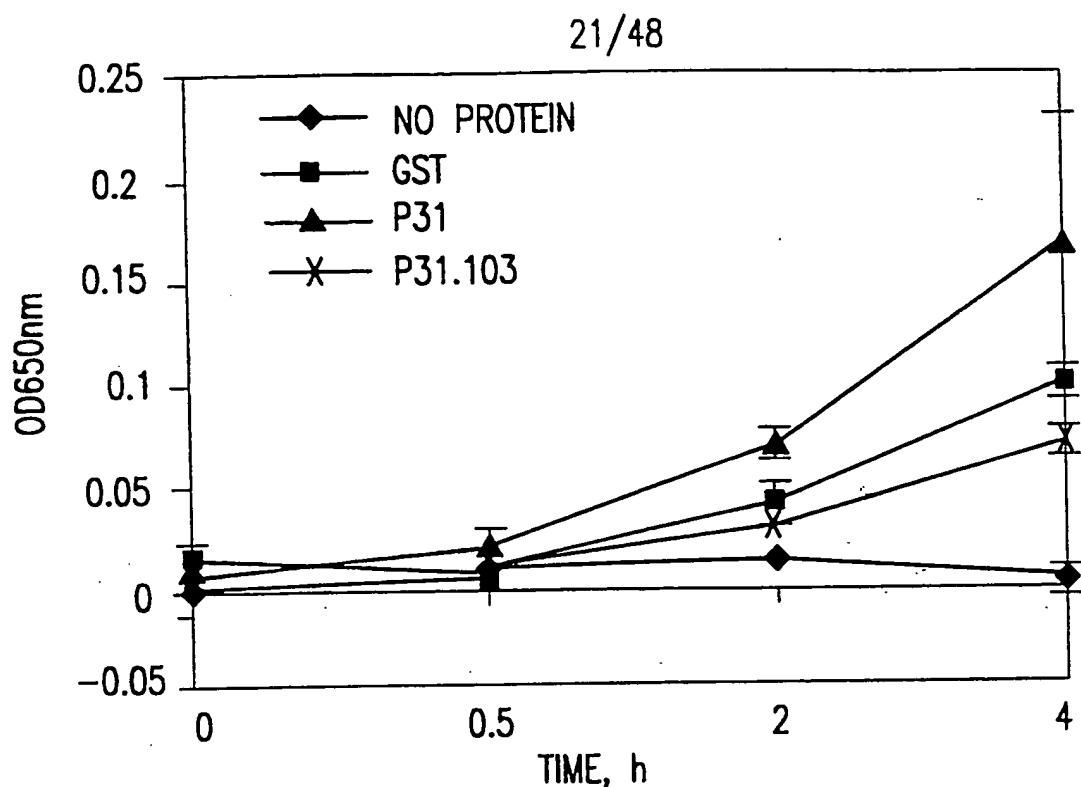


FIG.8A

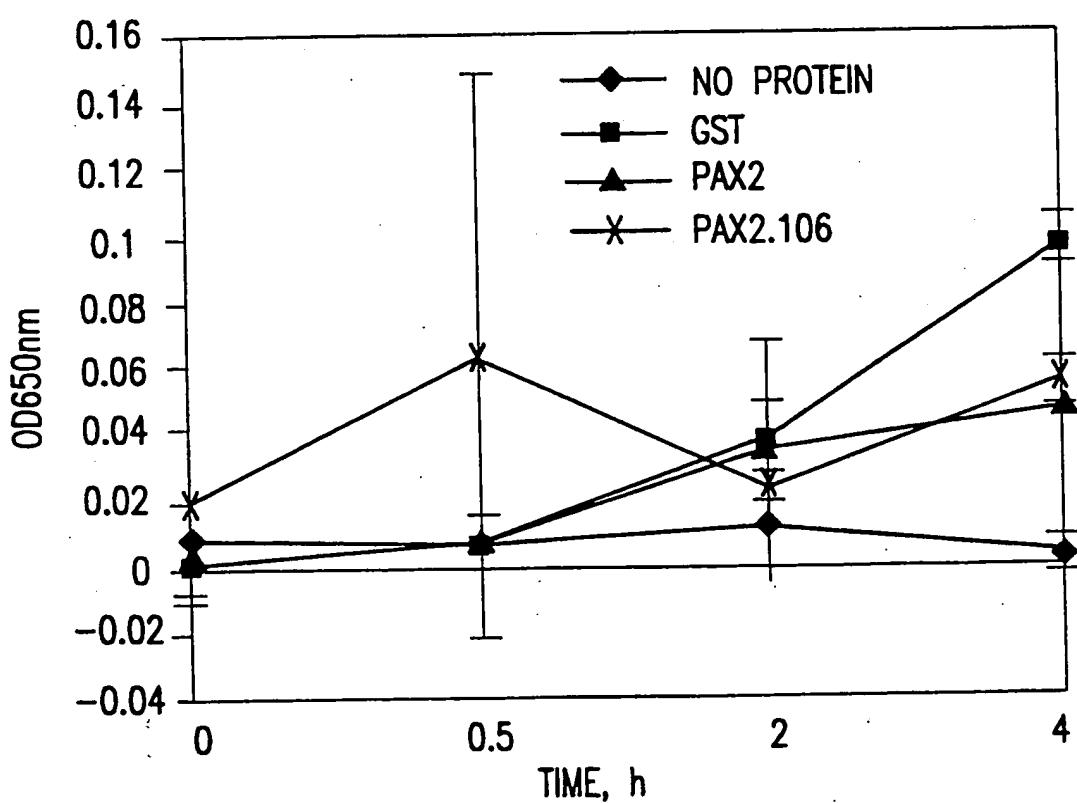


FIG.8B

22/48

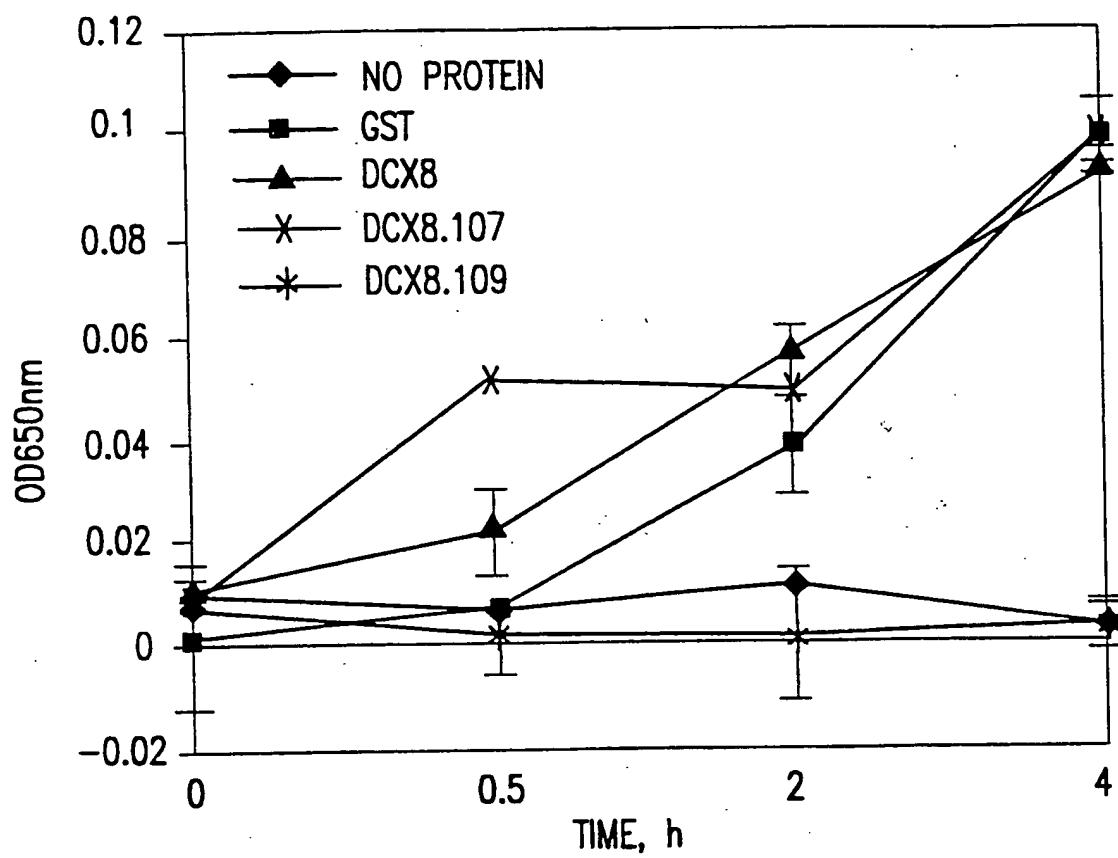


FIG.8C

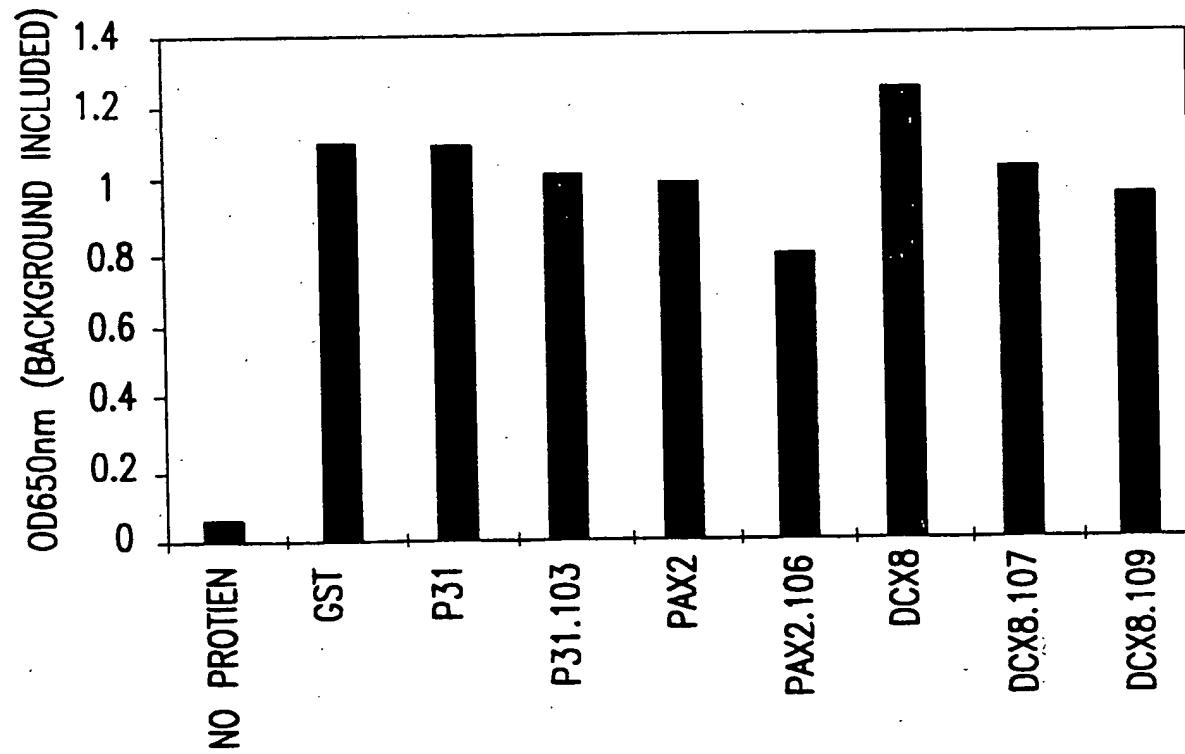


FIG.8D

23/48

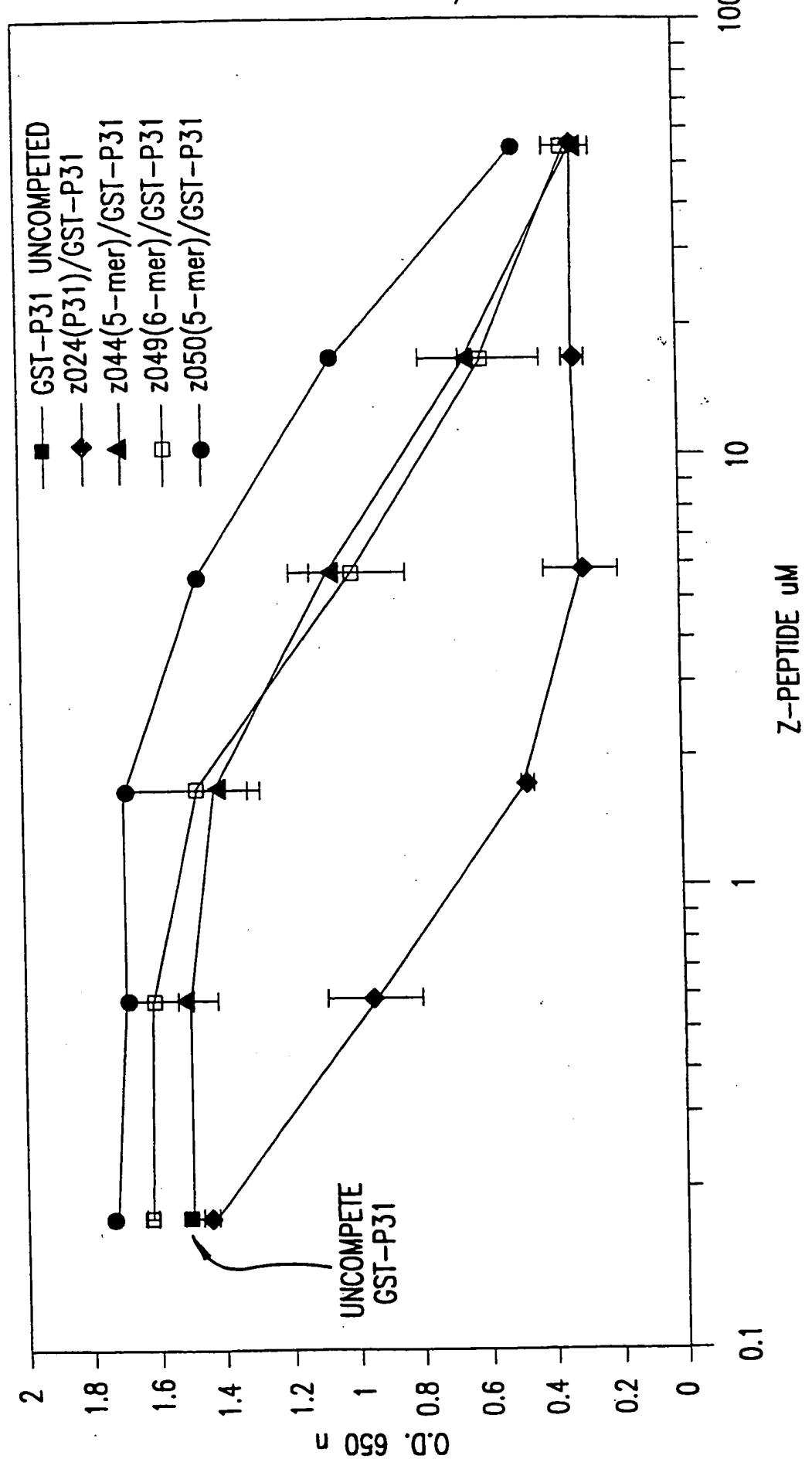


FIG. 9A

24/48

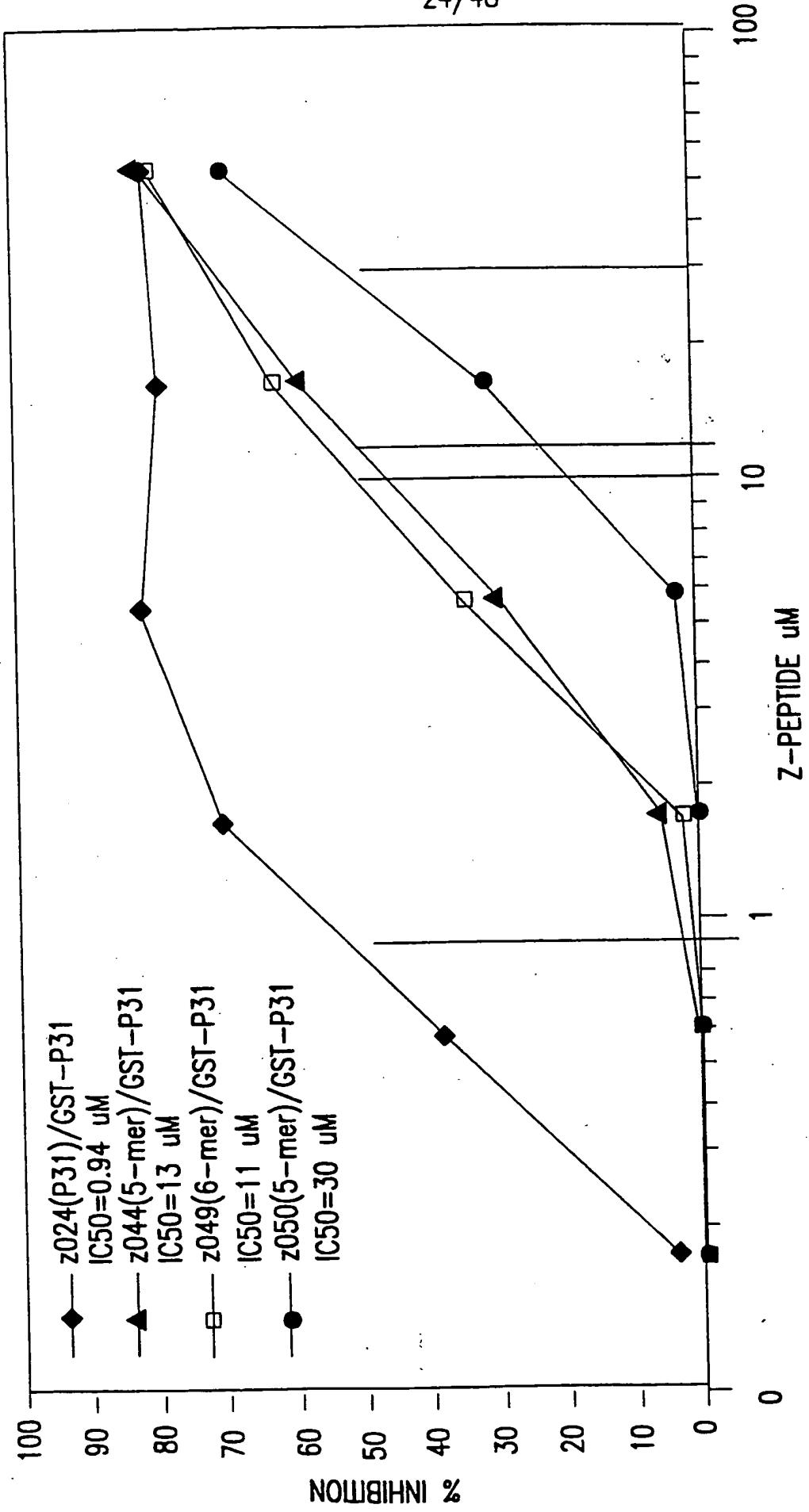


FIG. 9B

25/48

Peptide Name	Sequence	pI	IC <sub>50</sub>	GST/C2BBe1
ELAN024 (P31)	SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRRHPG	11.88	0.5-2.2	++
101	SARDSGPAEDGSRAVRLNGVENANTRKSSRSNPRGRRHPG			
102	DGSRAVRLNGVENANTRKSSRSNPRGRRHPG			
103	ENANTRKSSRSNPRGRRHPG			
110	ENANTRKSSR			
111	RKSSRSNPRG			
112	SNPRGRRKP			
119	TRKSSRSNPRG			
228	ZENANTRKSSRSNPRGRRHPG	12.28	0.5-1.7	
229	ZTRKSSRSNPRG	12.40	5.5-15	
230	ZENANTRKSSRSNPRG	11.81>50		
231	ZTRKSSRSNPRGRRHPG	12.70	0.6-3.2	
239	ZENANTRKSSR	10.89>50		
240	ZSNPRGRRHPG	12.40	5.9-29	
241	ZENANT	3.75	>50	
242	ZANTRKS	11.05	>50	
243	ZTRKSS	11.05	>50	
244	ZRKSSR	12.11	13->50	
245	ZKSSRSN	11.05	40-48	
246	ZSSRSNPG	10.04	>50	
247	ZRSNPRG	12.40	>50	
248	ZSNPRG	10.04	>50	
249	ZPRGRRH	12.40	11-20	
250	ZRRHPG	12.10	30	
251 (HepC core)	ZKSSRGN	12.40	>50	
252 (HepC p26664)	ZKTTSERSQPRGRRQPG	12.10	9.8	1.6
253	ZTrKSSrSNPGrGrrHPG			
254	ZTRKSSrSNPGRGrRHPG			
221 (HAX42)SDHALGTNLRSDDNAKEPGDYNCCGNGNSTGRKVFNRRRPSAIP	11.27	1.7		

FIG. 10A

Peptide Name	Sequence	pI	IC <sub>50</sub>	GST/C2BBe1
ELAN018 (PAX2)	1 STPPSREAYSRPYSVWSDSDTNAKHSSHNRRLRTRSPNG 104 STPPSREAYSRPYSVWSDSD 105 SRPYSVWSDSDTNAKHSSHNRRLRTRSPNG 106 TNAKHSSHNRRLRTRSPNG 107 TNAKHSSHNR 113 TNAKHSSHNR 114 SSSHNRRLRTR 115 RRLRTRRSRPN	10.88 0.6-0.9. 1	+++	++
Z32	ZTNAKHSSHNRRLRTRSPN	12.58	1.2	
Z33	ZTNAKHSSHNRRLRTR	12.58	1.6	
Z34	ZSSHNRRLRTRSPN	12.7		
Z35	ZSSHNRRLRTR	12.58	1.6	
Z26	ZSEANLDGRKSRYSSPRRNSSTRPRTSPNSVHARYPSTDHD	10.88	0.38 - 1.8	2.7
Z38	ZSRANTDGRKSRYSSPRRNSSTEPRLSPNSVHARYPSTDHD	10.88	7-8. 3	
Z55	ZTNAKHSSHNR	10.88	1.7	42
Z56	ZRRRLRTTRSPN	12.7		
Z57	ZRRRLRTTRSP	12.58		
Z58	ZRRRLRTTR	12.58		
Z59	ZrrrlTrTrSrPN	12.58		NOT DONE
Z73	ZASHNRRLRTR	1.5		5.5
Z74	ZSAHNRRRLRTR	6.2		
Z75	ZSSANRRLRTR	1.6		
Z76	ZSSHARRRLRTR	1.8		
Z77	ZSSHNARLRTR	3.9		5.2
Z78	ZSSHNRRAARTR	4.5		4.6
Z79	ZSSHNRRLATR	1.4		5.2
Z80	ZSSHNRRLRAR	3.4		2.2
Z81	ZSSHNRRLRTA	3.4		
Z82	ZSSHNRRLRPTA	3.4		
Z21 (HAX42)	ZSDHALG TGNLRS DNAKEP GDYNC GNGN STGRKV FNRRRPSA IPT	11.27	0.7	

26/48

FIG. 10B

27/48

Peptide Name	Sequence	PI	IC <sub>50</sub>	GST/C2BBe1
ELAN016 (SNi10) 116 117 118	1 10 20 30 40 RVGQCTDSVRRPWARSCAHQGGAGTRNSHGCITRPLRQASAH RVGQCTDSVRRPWARSCA VRRPWARSCAHQGGAGTRNS	10 19	0.22	++ - + +/-
Z17 Z16C23 Z36 Z37	2RVGQCTDSVRRPWARSCAH ZCGAGTRNSHGCITRPLRQASAH ZVRRPWARSCAHQGGAGTRNS ZCTDSVRRPWARSC	8.66 9.03 11.62 8.01	3.6 0.7 0.27 3	
Peptide Name	Sequence	PI	IC <sub>50</sub>	GST/C2BBe1
ELAN021 (HAX42) ELAN018 (PAX2) Z26 Z38 Z34 (PAX2 14mer)	1 10 20 30 40 SDHALGTNLRSNAKEPGDYNCCGGNSTGRKVFNRRRPSAIP STPPSREAYSRRPYSVDSDDSDTNAAKHSSHNRRLRTRSRPN ZSEANLDGRKSRYYSSPRNSSTRPRTSPNSVHARYPSTDHD ZSRANTDGRKSRYYSSPRNSSTEPRLSPNSVHARYPSTDHD ZSSHNRRLRTRSRPN	11 27 10.88 10.88 10.88 12.7	5.5 0.23 <0.2 <0.2 0.33	++ ++ ++

FIG. 10C

28/48

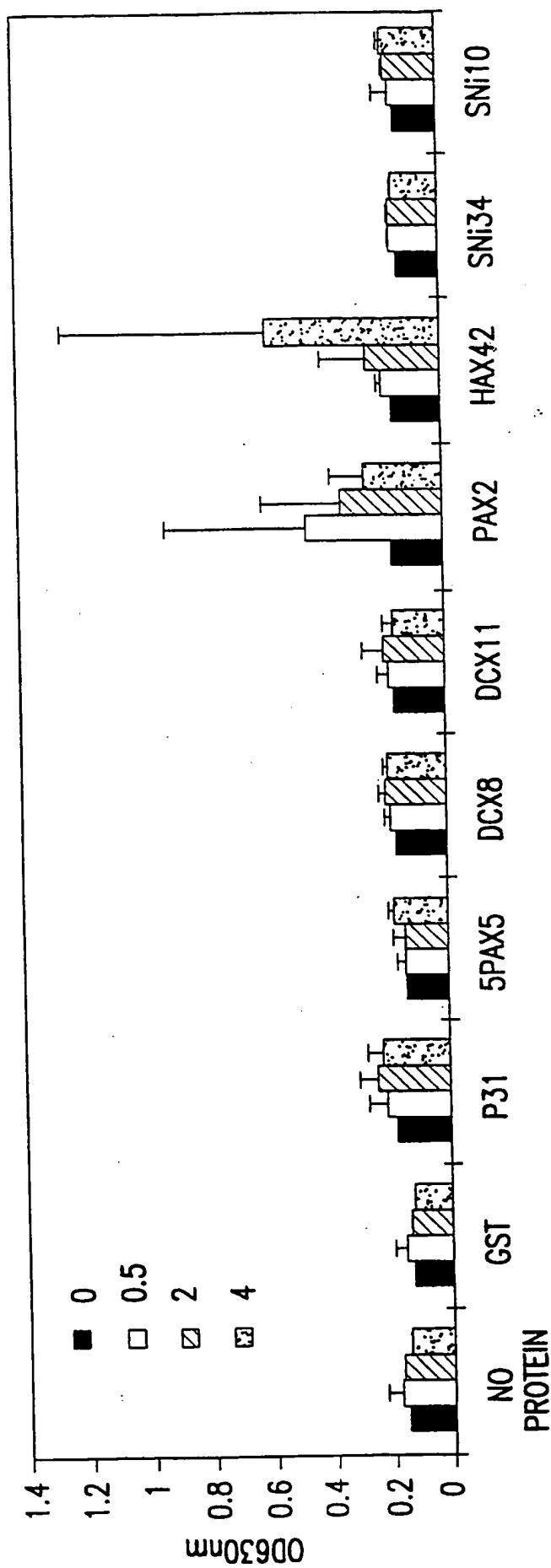


FIG. 11A

29/48

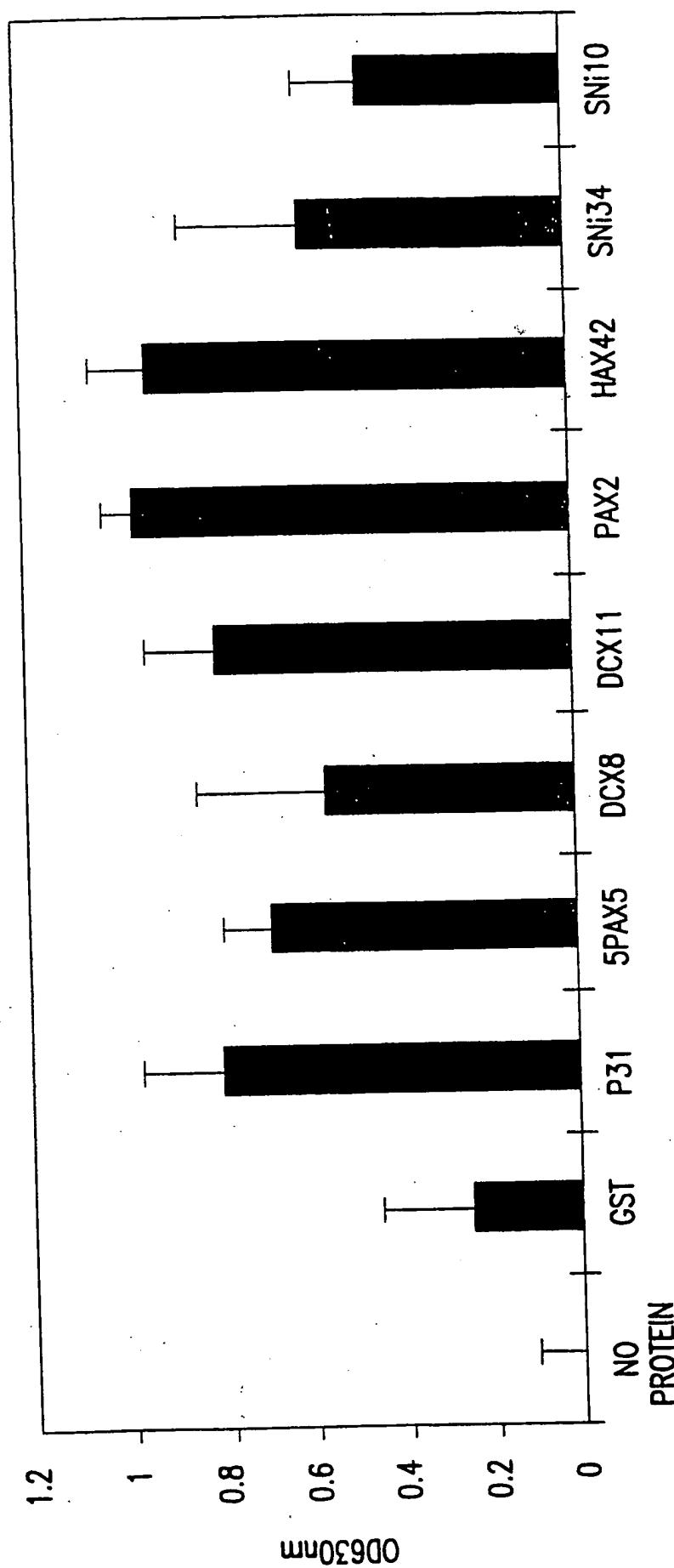


FIG. 11B

30/48

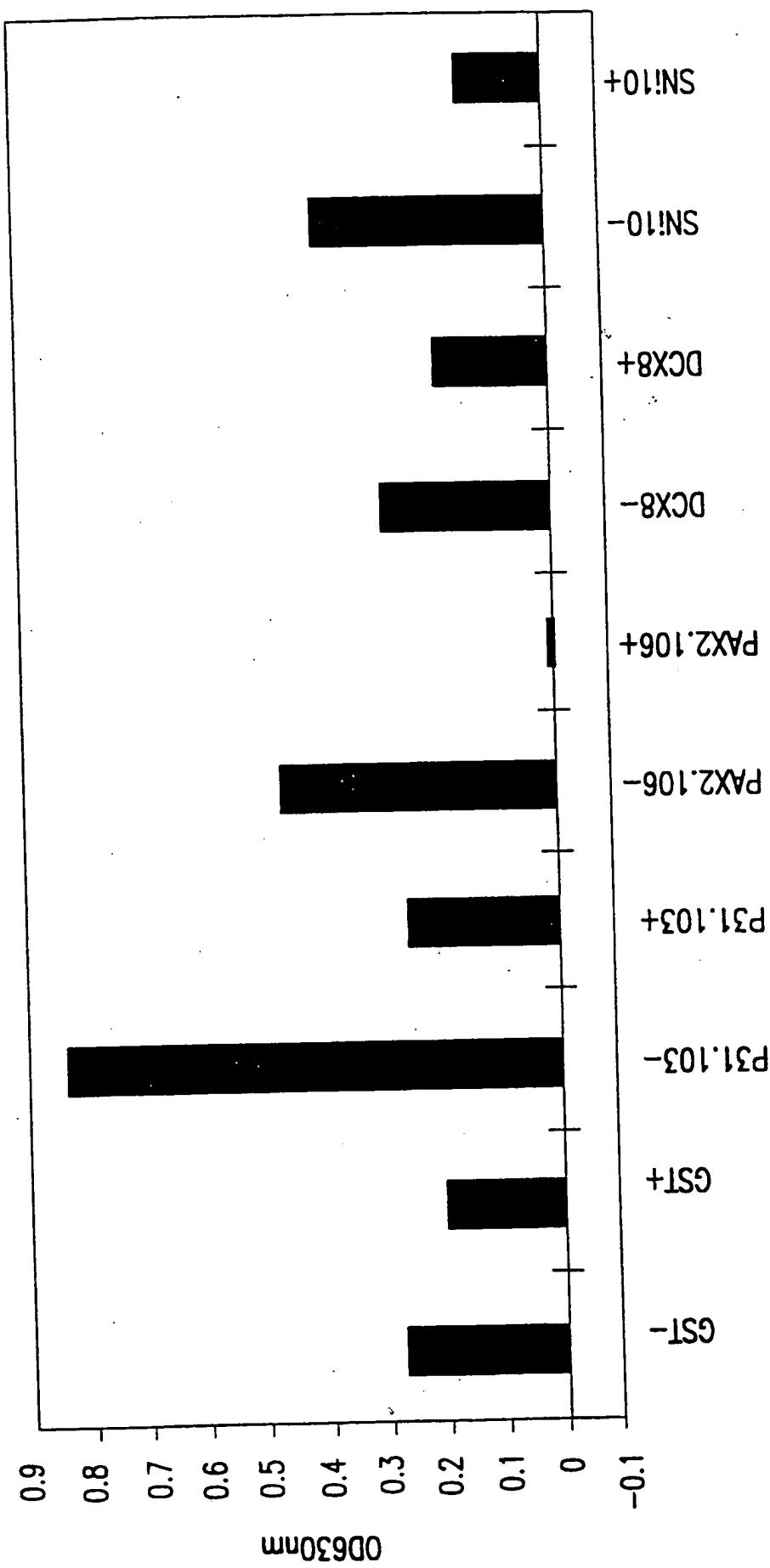


FIG. 12

31/48

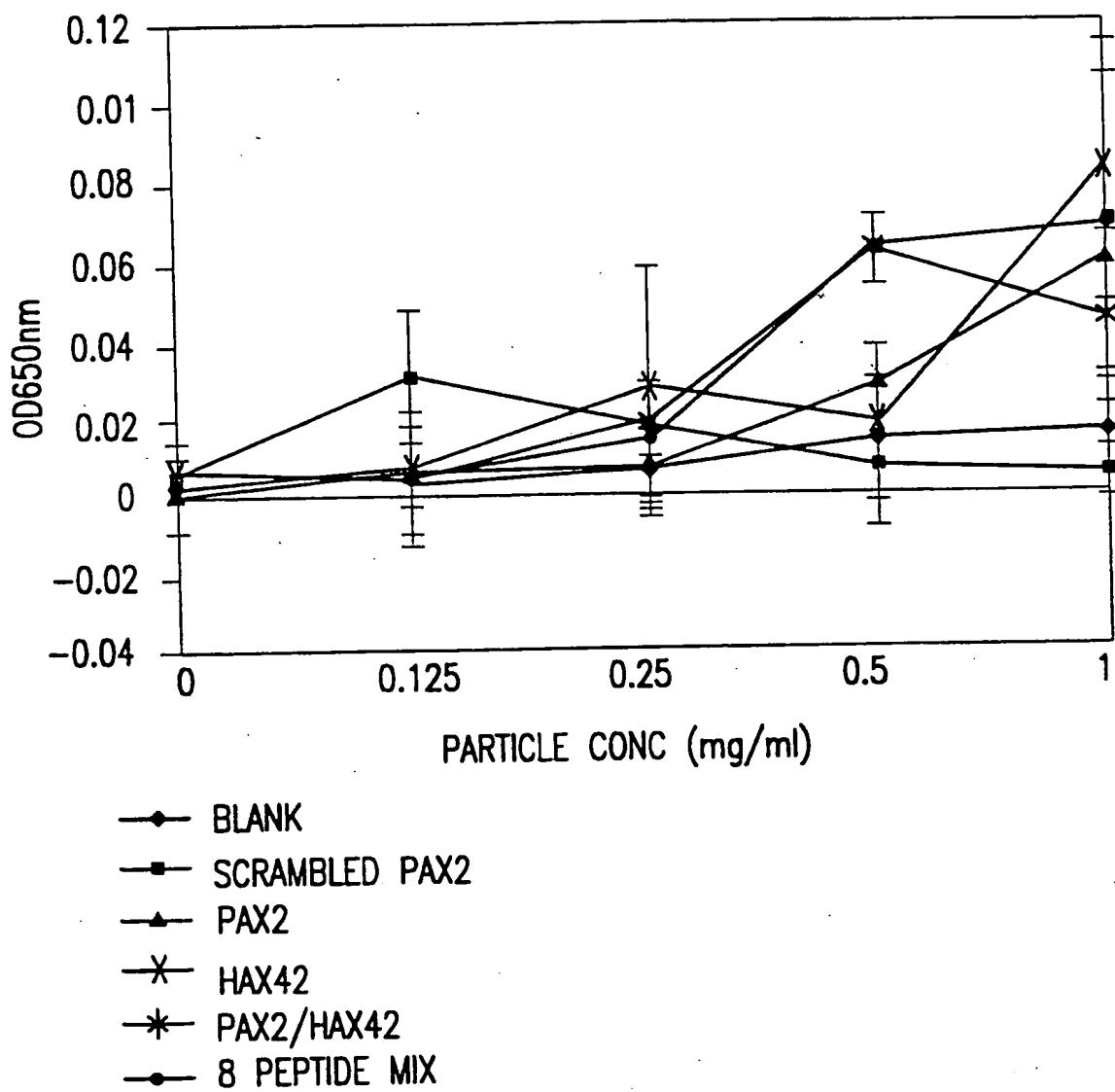


FIG.13A

32/48

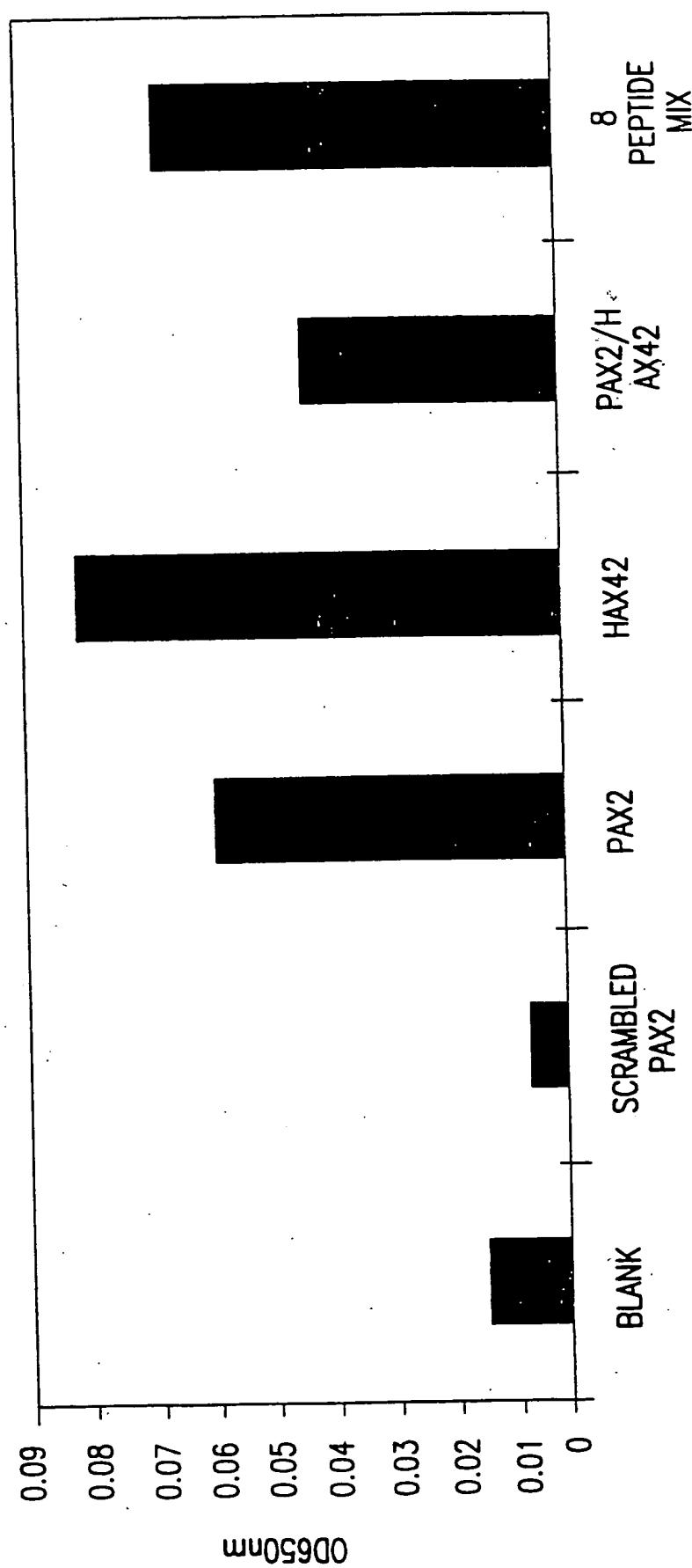


FIG. 13B

33/48

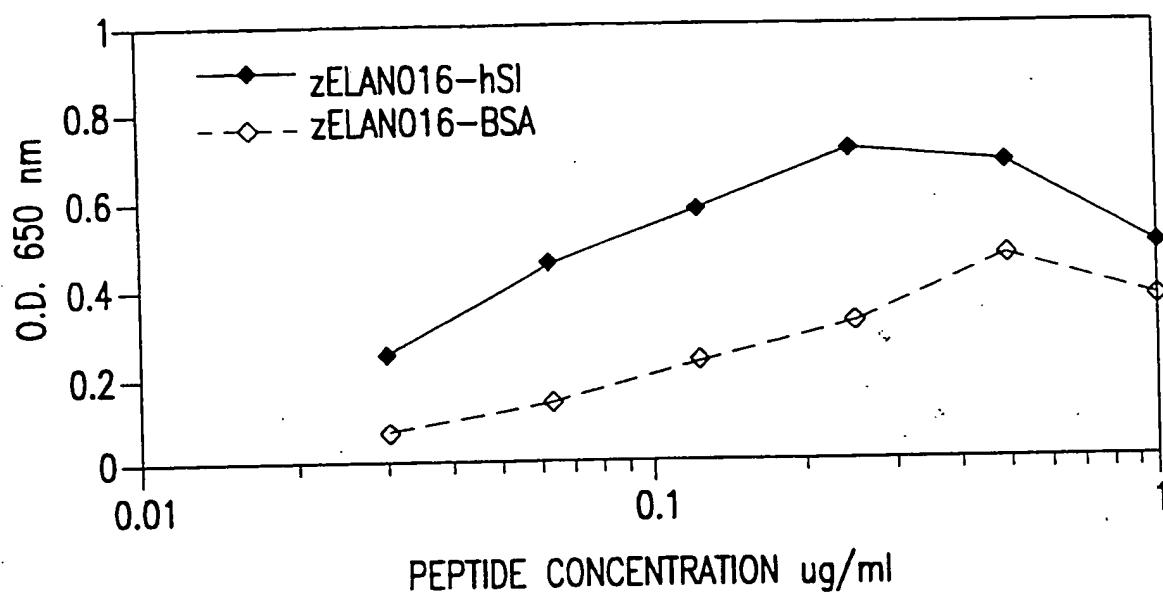


FIG. 14A

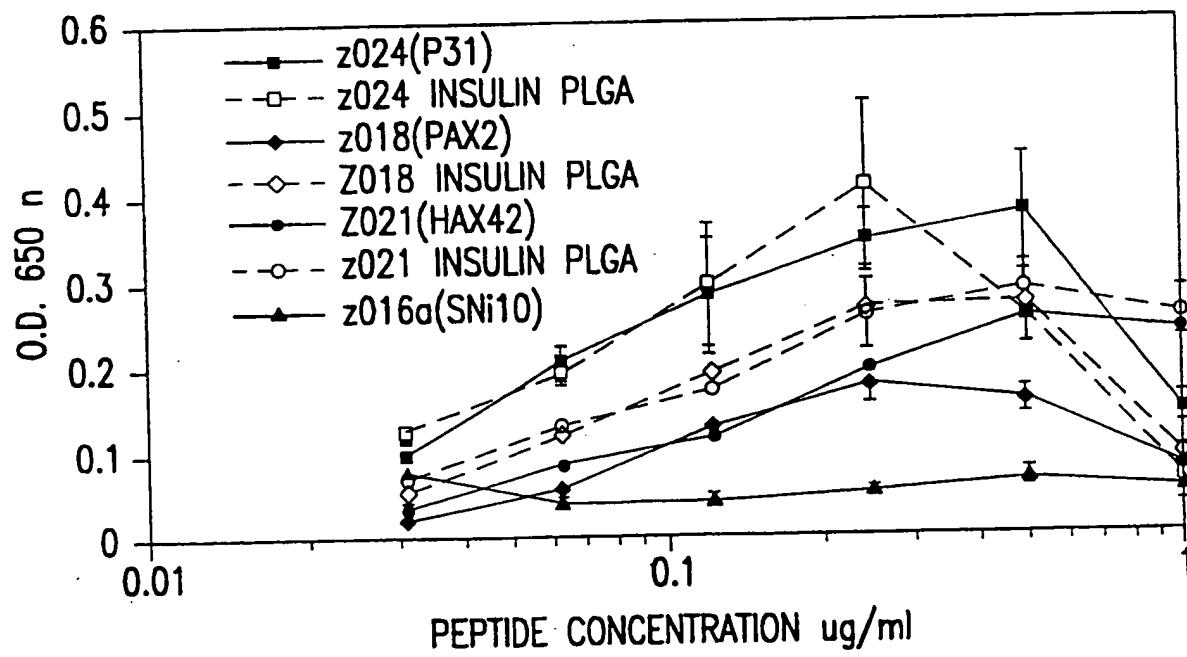


FIG. 14B

34/48

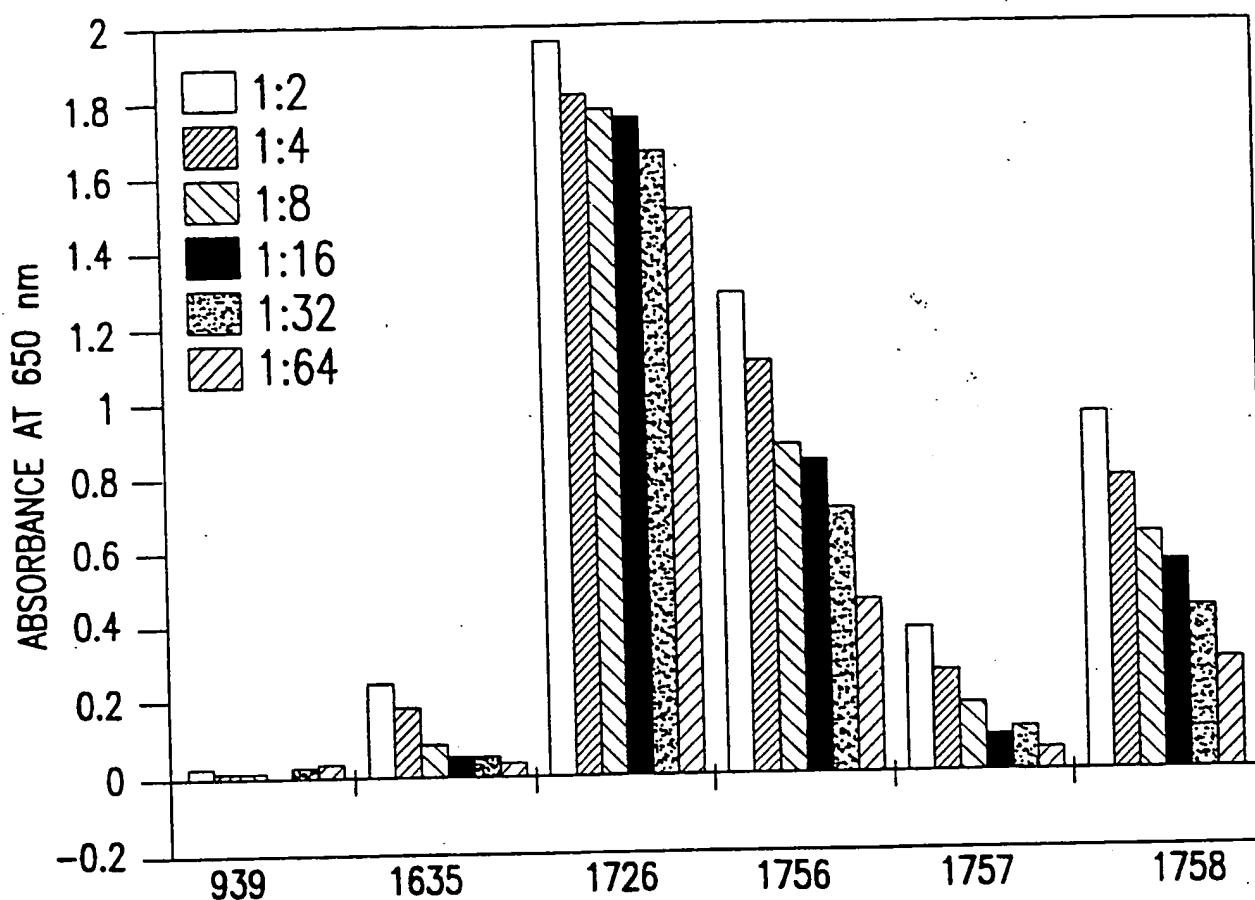


FIG. 15A

35/48

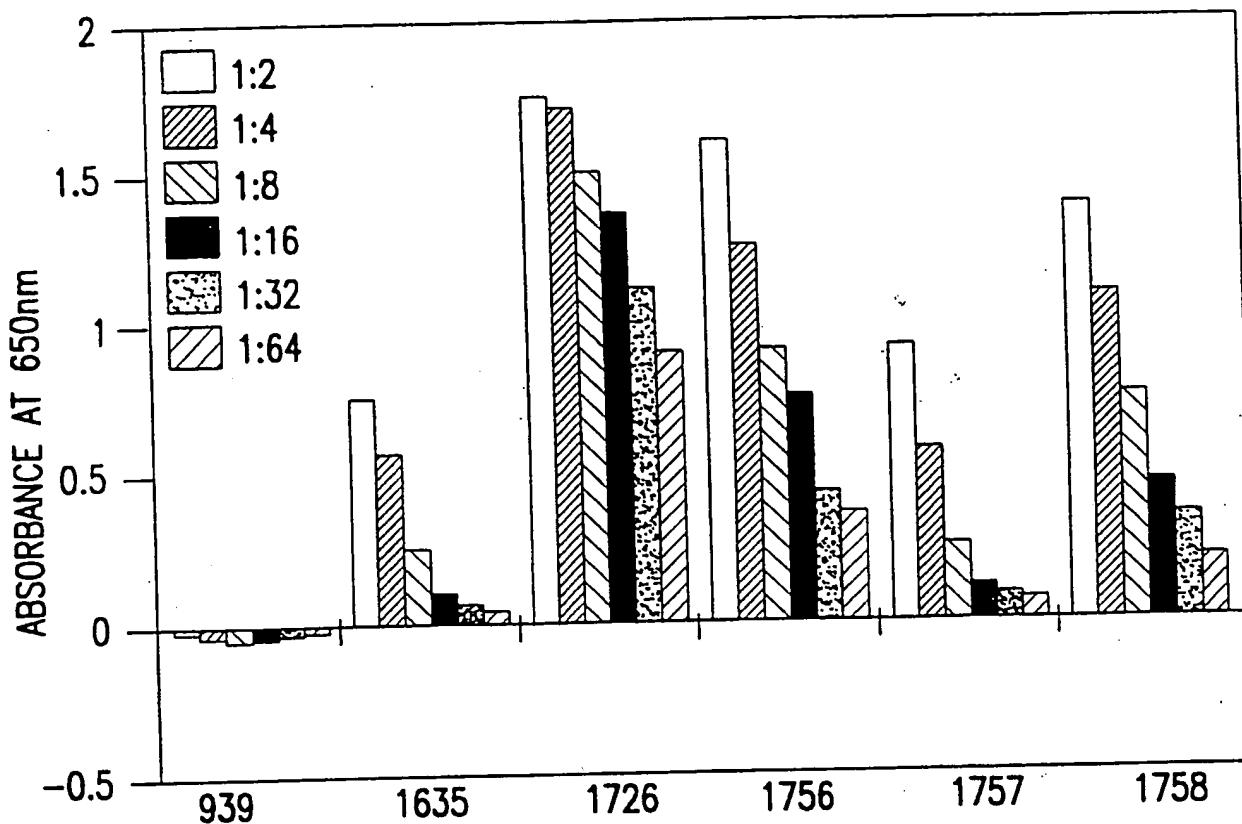


FIG. 15B

36/48

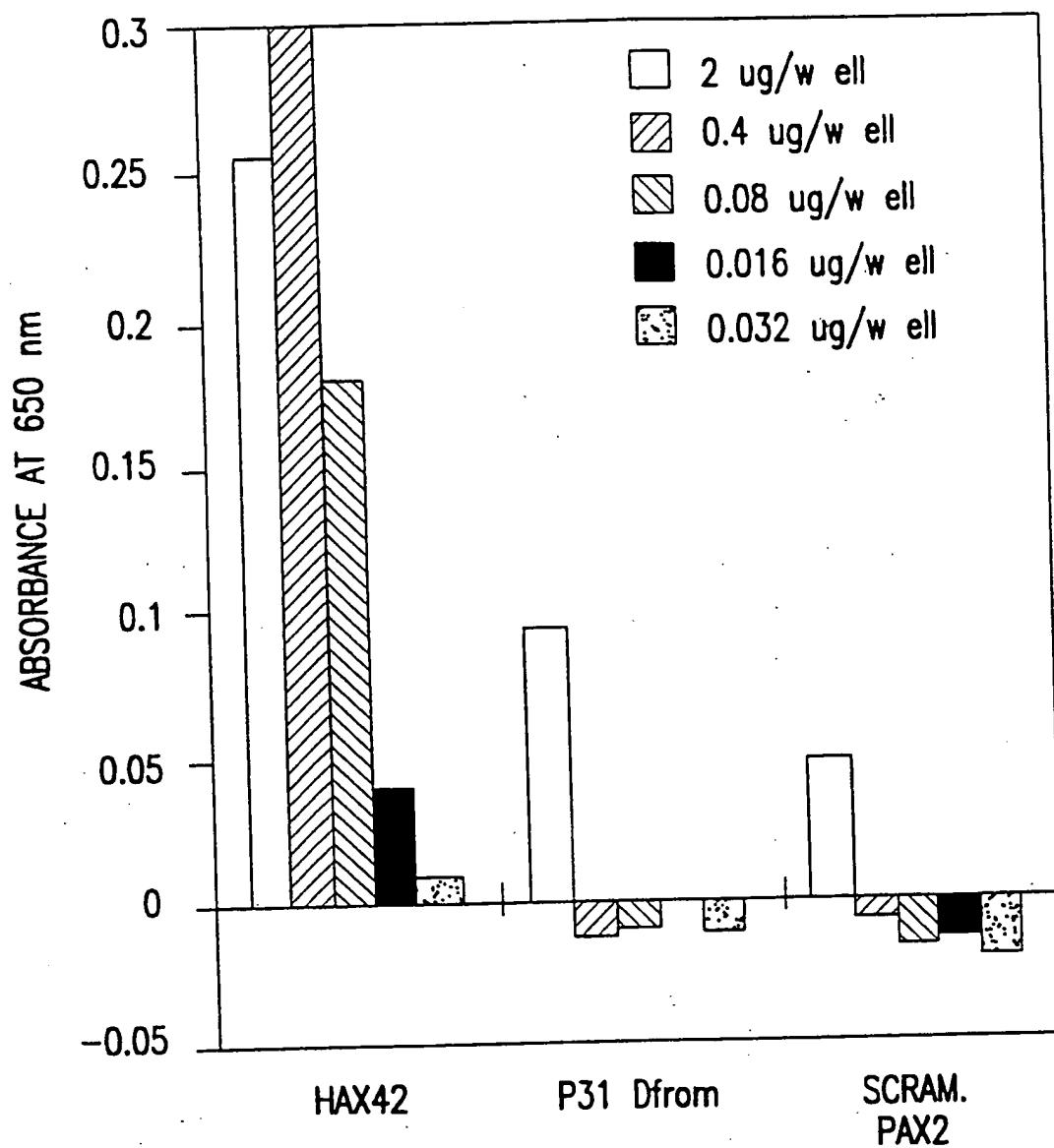


FIG.16A

37/48

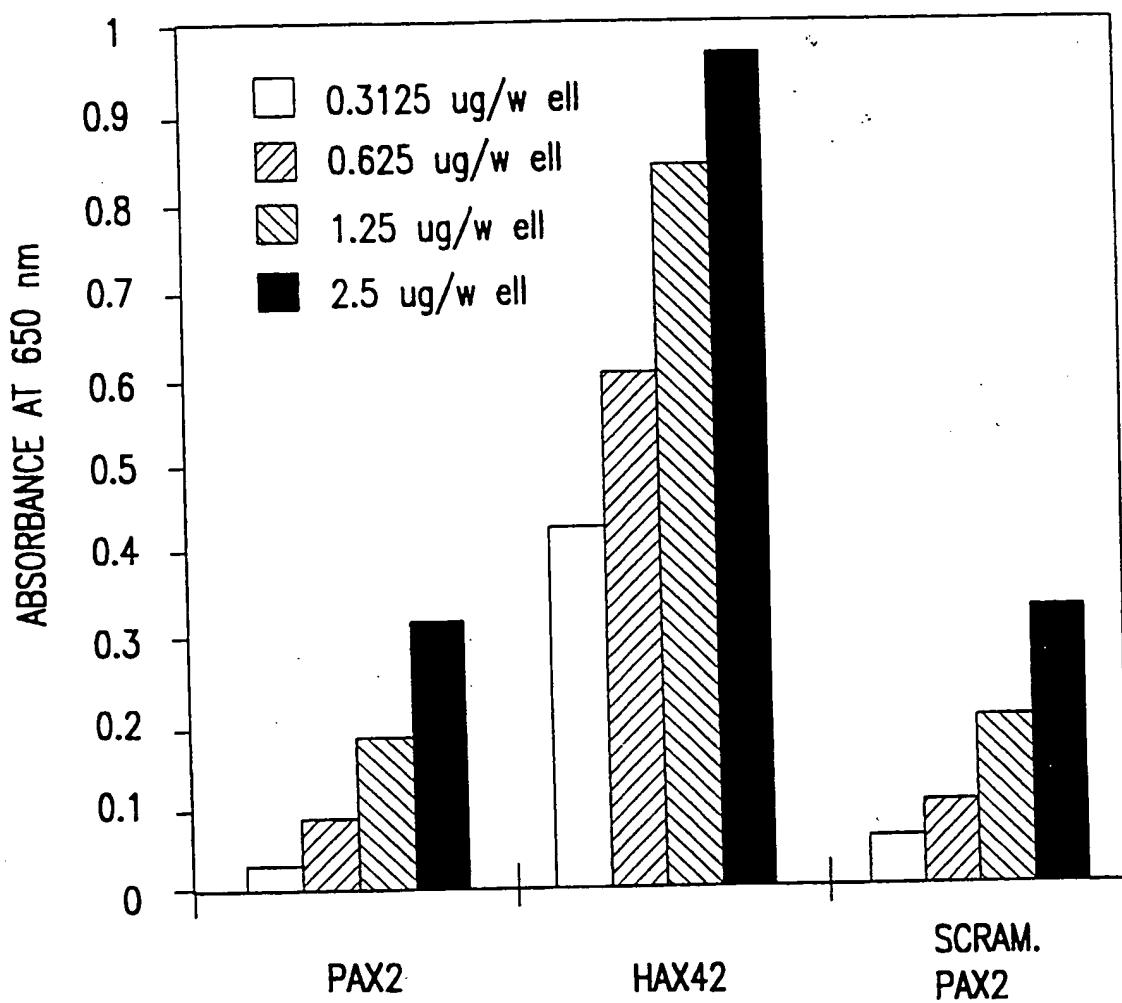


FIG.16B

38/48

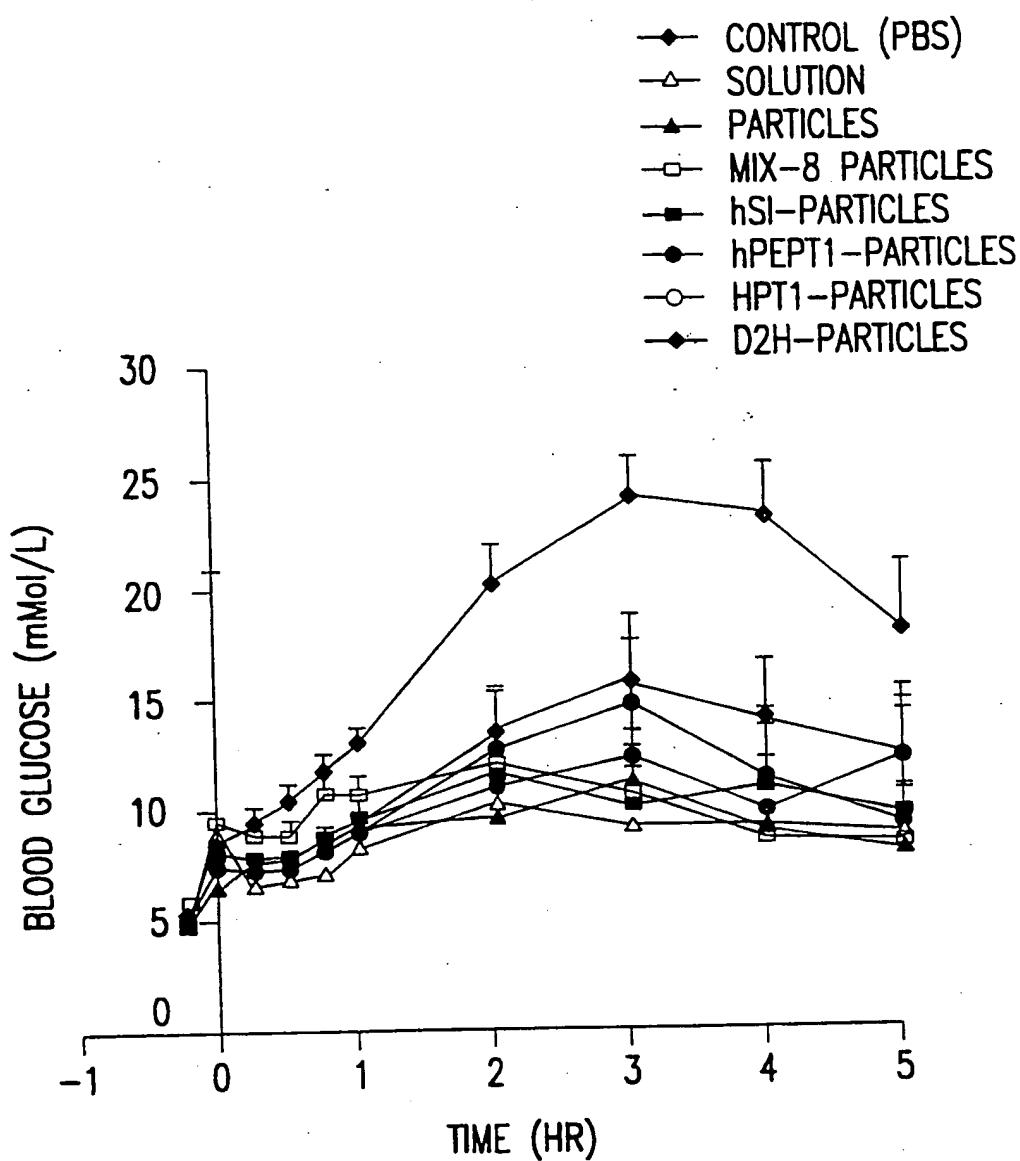


FIG. 17A

39/48

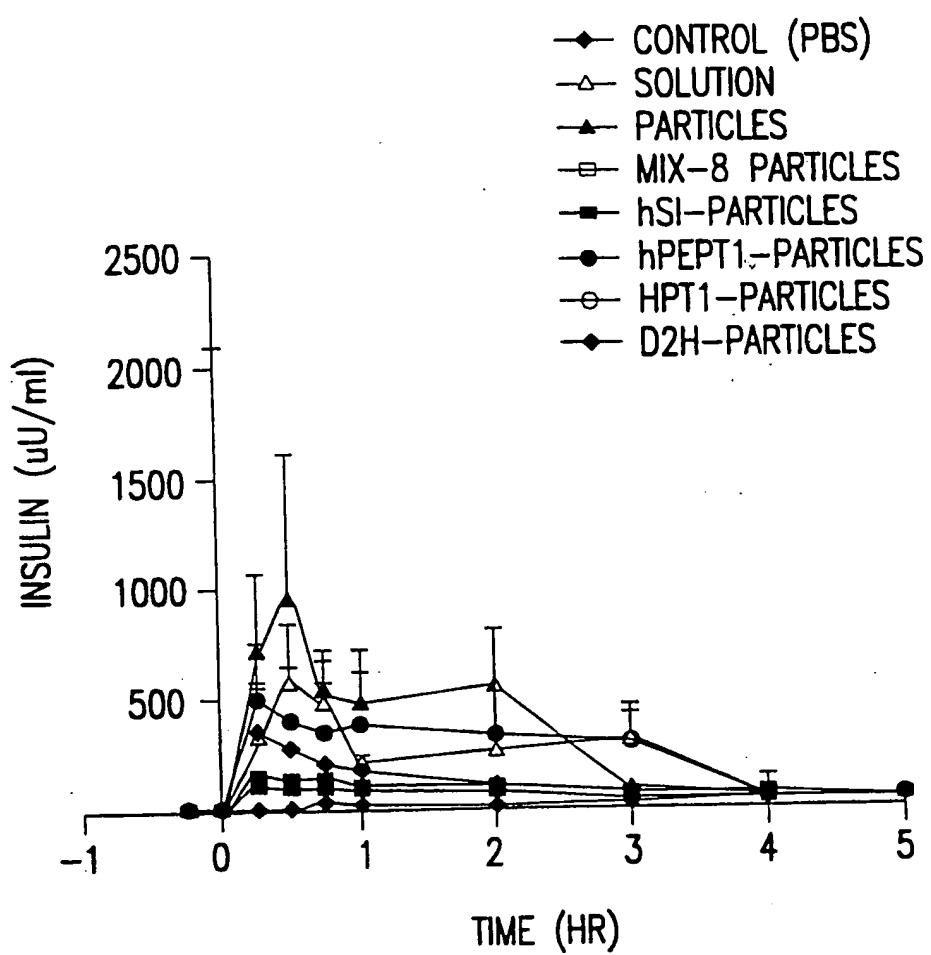


FIG. 17B

40/48

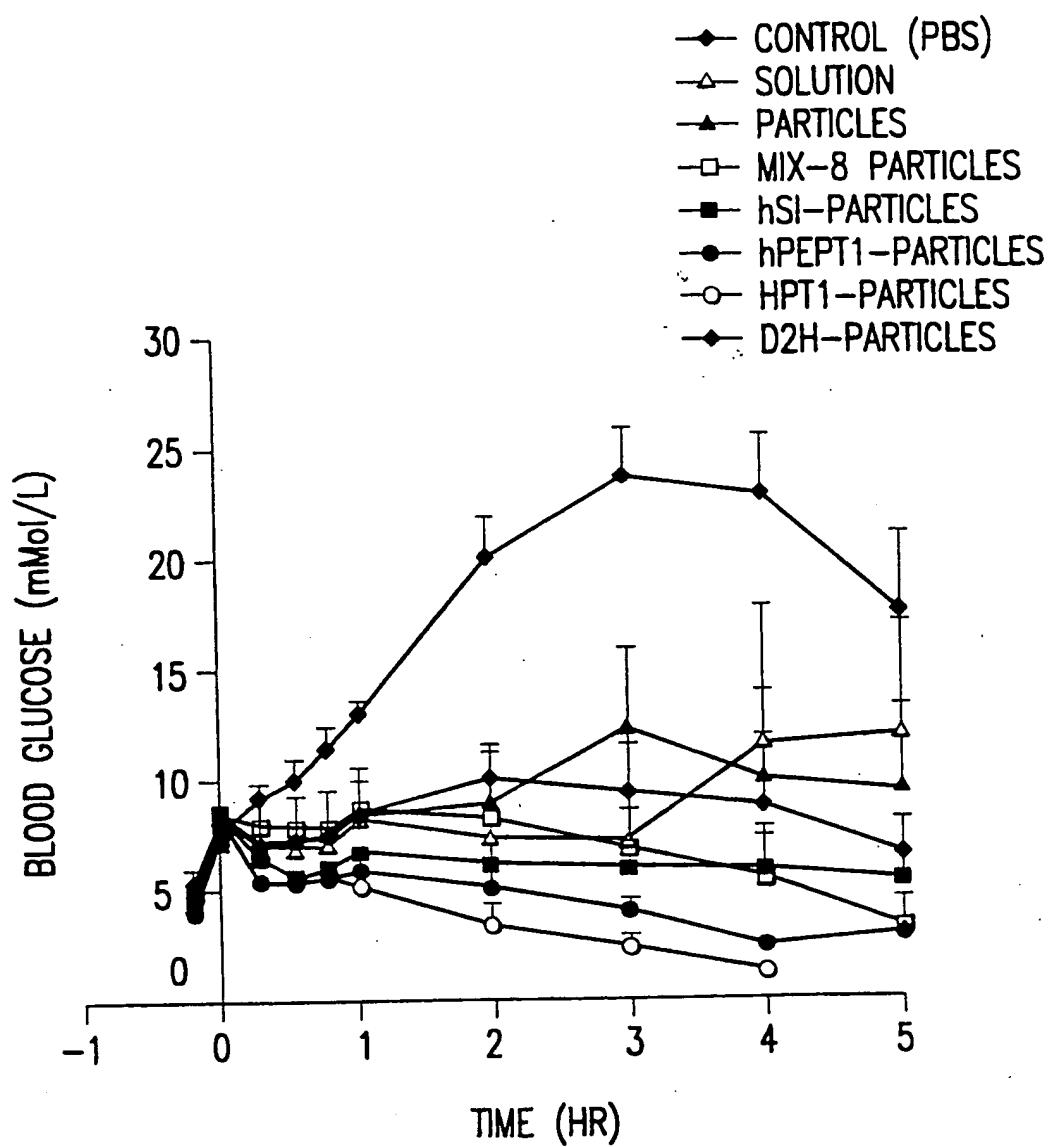


FIG. 18A

41/48

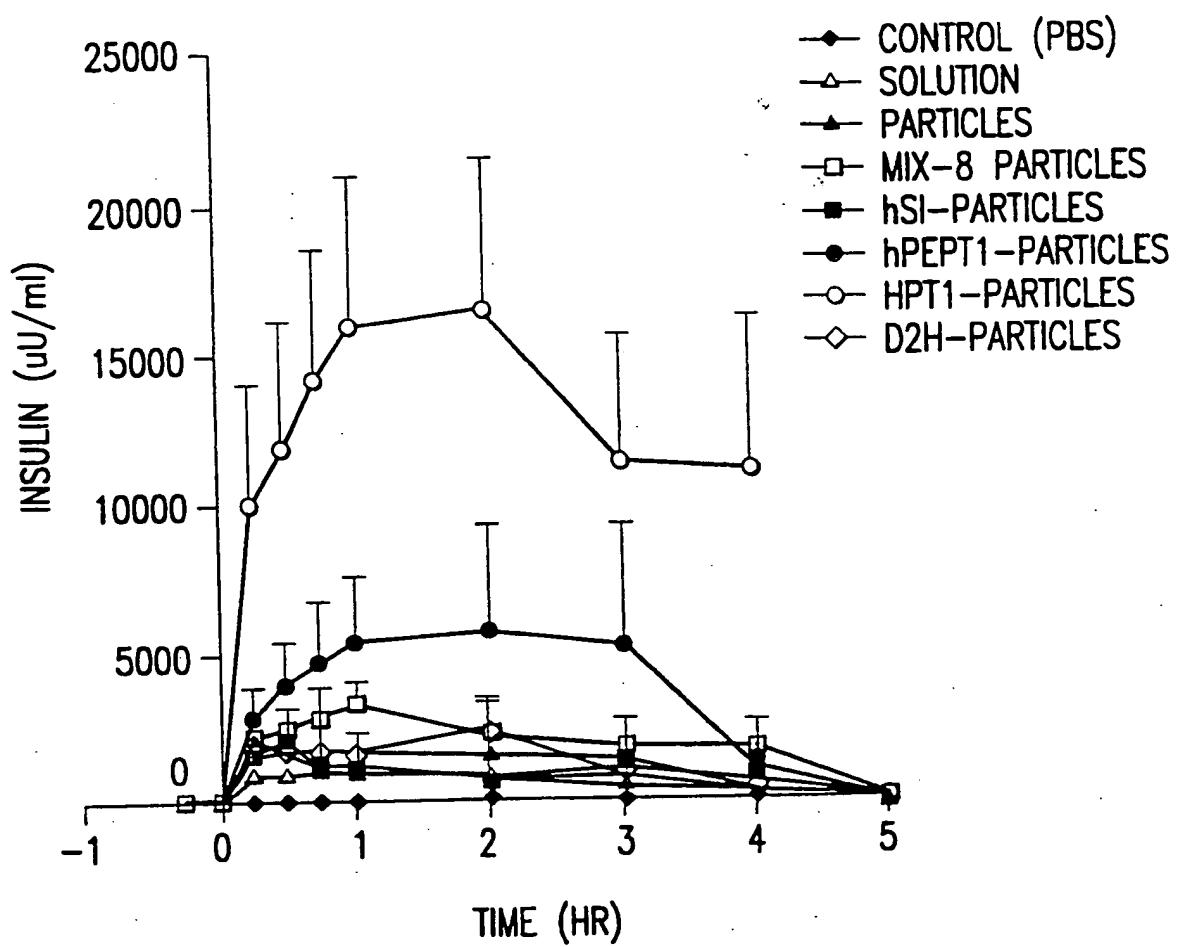


FIG.18B

42/48

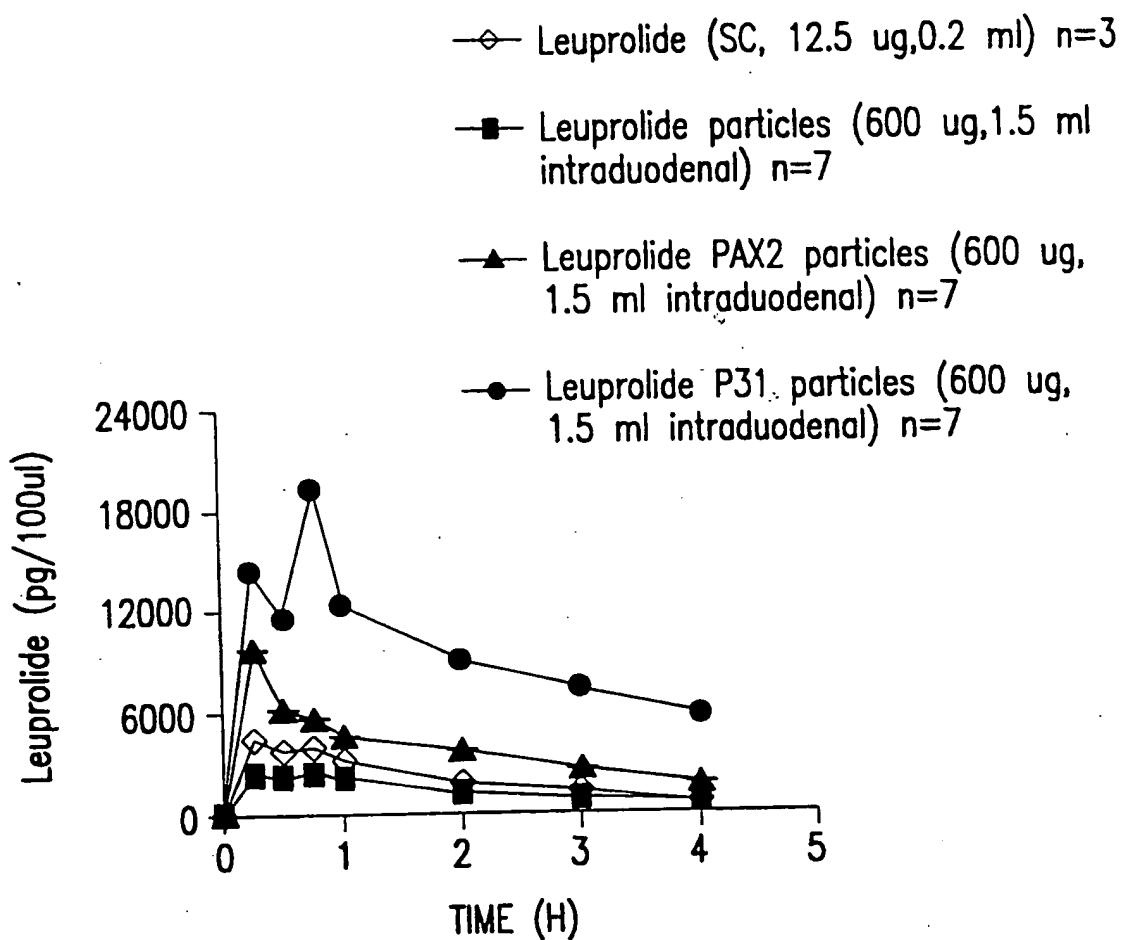


FIG. 19

P31 AA SEQ. POSITION	KNOWN PROTEIN	HOMOLOGOUS SEQ. POSITION
12-34	FASCICULIN 2	10-32
4-12	MESENTERICOPEPTIDASE	54-62
15-31		175-191
26-39	CORE PROTEIN (HEPATITIS C VIRUS)	5-18
26-39		11-24
26-39		21-34
26-39		38-51
23-30		39-55
25-39		41-55
26-39		51-64
16-39	PT-NANBH POLYPYPROTEIN N-TERMINUS	51-64
28-40	AL2 PROTEIN (CAENORHABDITISELEGANS)	70-82
26-38	CAPSID PROTEIN (HEPATITIS C VIRUS TYPE 3g)	48-60
26-39	GENOME POLYPYPROTEIN (HEPATITIS C VIRUS)	57-70

FIG.20

44/48

DCX8AA SEQ. POSITION	KNOWN PROTEIN	HOMOLOGOUS SEQ. POSITION
20-27	ENDO-1,4-BETA-D-GLUCANASE	78-85
30-37		221-228
21-34	P-HYDROXYBENZOATE HYDROXYLASE	285-298
5-15		54-64
7-21	CYTOCHROME	50-64
7-21	CYTOCHROME C3	50-64
	TRIMETHYLARNINE DEHYDROGENASE	208-219
32-43		396-407
30-37	Gag-JunD FUSION PROTEIN	24-31
26-30		16-20
23-44	SECRETIN PRECURSOR, N-PROSECRETIN, SECRETIN AMIDE	18-39
33-44	T-CELL RECEPTOR V BETA CHAIN	15-26
27-33		3-9
23-44	SECRETIN PRECURSOR PIR	18-39
31-44	HYPOTHETICAL PROTEIN V (SYNECHOCYSTIS)	275-288
24-30		251-257
23-43	PUTATIVE RNA BINDING PROTEIN	230-250
28-40	Mu SON OF SEVENLESS 1	1-13
24-35	NEUROPEPTIDE PRECURSOR	80-91
29-43		5-19
23-43	RNA-BINDING PROTEIN (MACACAFASCICULARIS)	230-250
23-43	RNA-BINDING PROTEIN (HOMOSAPIENS)	230-250
23-43	AUTOSOMAL GENE-AZOOSPERMIA FACTOR	230-250
25-38	COLLAGEN	25-28
24-35		4-15
29-41	PROBABLE CELL GROWTH REGULATOR	306-318
24-35	RIBOSOMAL PROTEIN S2	24-35
T6-39		182-185
24-44	CAENORHABDITIS ELEGANS	296-316
23-34	pid:e208155 (HOMO SAPIENS)	61-72
36-43		116-123

FIG.21A

45/48

DCX8A SEQ. POSITION	KNOWN PROTEIN	HOMOLOGOUS SEQ. POSITION
24-38	XYLULOSE KINASE	16-30
24-39	CAENORHABDITIS ELEGANS	57-72
26-42		65-81
27-33	HYPOTHETICAL PROTEIN-PHAGE BZ13	22-28
35-39		31-35
30-42	CEREBELLIN-LIKE GLYCOPROTEIN	2-14
8-22	DNA PRIMASE	170-184
2-7		76-81
5-21	COAT PROTEIN (BEAN COMMON MOSAIC VIRUS)	12-28
5-21	COAT PROTEIN (BEAN COMMON MOSAIC VIRUS)	33-49
5-21		19-35
5-21	POLYPROTEIN (BEAN COMMON MOSAIC VIRUS)	215-231
5-21		39-55
5-21	Nib PROTEINCOAT PROTEIN (COWPEA APHID-BOME MOSAIC VIRUS)	92-108
2-13	MHC CLASS 1 PIP1 (PITHECIA)	111-122
14-22		236-334
3-19	TALIN (CAENORHABDITIS ELEGANS)	1538-1554
2-9	ACETAMIDASE PIR	359-366
9-20		483-494
10-16	RHIZOBIONS ETLI STRAIN	134-140
17-30		173-186
31-39		200-208
2-11	NEUROTOXIN 1 (TOXIN B) A. STOKESI	7-16
12-33		26-47
21-27	SUID HERPES VIRUS 1 EARLY PROTEIN	425-432
30-43		51-64
13-42	RICE cDNA PARTIAL SEQUENCE	50-151
8-15	FUSION PROTEIN	24-31
4-8		16-20
1-22	SECRETIN PRECURSOR, N-PROSECRETIN, SECRETIN-AMIDE	18-39
11-22	T-CELL RECEPTOR V BETA CHAIN	15-26
5-11		3-9
9-22	HYPOTHETICAL PROTEIN	275-288
2-8		251-257

46/48

DCX8A SEQ. POSITION	KNOWN PROTEIN	HOMOLOGOUS SEQ. POSITION
1-21	PUTATIVE RNA BINDING PROTEIN	230-250
6-18	HYPOTHETICAL PROTEIN-MOUSE PIR	1-13
2-13	NEUROPEPTIDE PRECURSOR	80-91
7-21	orf3-HUMAN	5-19
1-21	RNA-BINDING PROTEIN	230-250
13-16	COLLAGEN	25-28
7-19	PROBABLE CELL GROWTH OR DIFFERENTIATION REGULATOR	306-318
2-13	RIBOSOMAL PROTEIN S2	14-25
14-17		182-185
2-22	CAENORHABDITIS ELEGANS	296-316
1-12	HOMOSAPIENS	61-72
14-21		116-123
2-16	XYLULOSE KINASE	16-30
8-15	T CELL RECEPTOR DELTA CHAIN	55-62
5-8		12-15
8-17	SEQ. 43 FROM PATENT US	12-21

FIG.21C

47/48

DAB10 AA SEQ. POSITION	KNOWN PROTEIN	HOMOLOGOUS SEQ. POSITION
13-34	1,3-BETA-GLUCANASE	231-252
3-11	PHOTOSYNTHETIC REACTION CENTER	20-28
16-27		128-139
28-35	MYB PROTO-ONCOGENE PROTEIN	131-138
5-18		32-45
23-36	LYSOZYME MUTANT	130-143
28-35	LIPASE	400-407
3-15		159-171
3-37	TRYPSIN	169-203
13-34	1,3-1,4-BETA-GLUCANASE	232-253
4-10	LACTATE DEHYDROGENASE	190-196
11-7		244-250
4-10	APO-LACTATE DEHYDROGENASE	190-196
11-17		244-250
4-10	LACTATE DEHYDROGENASE	191-197
11-17		245-251
16-26	OVOTRANSFERRIN	240-250
23-36	GENOME POLYPROTEIN MATRIX PROTEIN	1022-1035
14-20	ROUS SARCOMA VIRUS	43-49
2-12		13-23
14-20	HYPOTHETICAL PROTEIN-AVIAN LEUKOSIS VIRUS	43-49
4-20	T CELL RECEPTOR DELTA CHAIN VARIABLE REGION	1-4
14-18		12-16
2-12	GAG POLYPROTEIN-AVIAN ENDOGENOUS VIRUS RAV-0	139-149
14-20		169-175
	p19 PROTEIN-AVIAN ERYTHROBLASTOSIS VIRUS	189-199
14-20		219-225
7-19	ALI PROTEIN-POTATO YELLOW MOSAIC VIRUS	222-234
3-22	ENDO-1,4-BETA GLUCANASE	186-205
6-18	I a PROTEIN-BROME MOSAIC VIRUS	430-442
2-12	GAG POLYPROTEIN-FUJINAMI SARCOMA VIRUS	186-196
14-22		216-222
2-12	GAG PROTEIN-ROUS SARCOMA VIRUS	190-200
14-20		220-226
1-12	CORTICOTROPIN-LIKE INTERMEDIATE LOBE PEPTIDE	7-18
1-22	GENE PRODUCT (CAENORHABDITIS ELEGANS)	4-25
31-37	T CELL RECEPTOR DELTA CHAIN	56-62
26-39		12-15
26-37	LYSOZYME MUTANT	133-144

FIG.22

## 48/48

ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro 1 5 10 15	48
ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG Thr Arg Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu 20 25 30	96
TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu 35 40 45	144
GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys 50 55 60	192
TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn 65 70 75 80	240
ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu 85 90 95	288
GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser 100 105 110	336
AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu 115 120 125	384
ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT Met Leu Lys Met Phe Gln Asp Arg Leu Cys His Lys Thr Tyr Leu Asn 130 135 140	432
GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp 145 150 155 160	480
GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu 165 170 175	528
GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr 180 185 190	576
TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala 195 200 205	624
ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG GTT CCG CGT Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg 210 215 220	672
GGA TCC CCA GGA ATT CCC GGG TCG ACT CGA GCG GCC GCA TCG TGA Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser 225 230 235	717